Thesis for a Ph. D. Degree

Improved retrieval of cloud parameters for IASI 1D-Var data assimilation and its impact on NWP model forecast

IASI 일차변분법 자료동화 시스템 내 구름 변수 산출 알고리즘 개선과 수치예보 정확도에 미치는 영향

August 2020

School of Earth and Environmental Sciences

Graduate School

Seoul National University

Ahreum Lee

Improved retrieval of cloud parameters for IASI 1D-Var data assimilation and its impact on NWP model forecast

By

Ahreum Lee

A Dissertation submitted to the Faculty of the Graduate School of the Seoul National University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Degree Awarded:

August 2020

Advisory committee:

Professor Seok-Woo Son, Chair

Professor Byung-Ju Sohn, Advisor

Professor Myung-Hwan Ahn

Doctor Yoonjae Kim

Doctor Hyoung-Wook Chun

Abstract

The Unified Model (UM) data assimilation system incorporates a one-dimensional variational (1D-Var) analysis of cloud variables for hyperspectral infrared sounders that allows the assimilation of radiances in cloudy areas. For the Infrared Atmospheric Sounding Interferometer (IASI) radiance assimilation in the UM, a first guess pair of cloud top pressure (CTP) and cloud fraction (CF) is estimated using the minimum residual (MR) method, which simultaneously obtains CTP and CF by minimizing radiances difference between observation and model simulation. In this study, specific pairs of CTP and CF yielding the smallest 1D-Var temperature and humidity analysis error were found from the ECMWF short-range forecast based IASI simulated radiances and background states, and defined as ‘optimum’ cloud parameters. Compared to the optimum results, it is noted that the MR method tends to overestimate cloud top height while underestimating cloud fraction. This fact necessitates an improved cloud retrieval for better 1D-Var analysis performance.

An Artificial Neural Network (ANN) approach was taken to estimate CTP as close as possible to the optimum value, based on the hypothesis that CTP and CF closer to the optimum values will bring in better 1D-Var results. The ANN-based cloud retrievals indicated that CTP and CF biases and root mean square errors against the optimum values shown in the MR method are much reduced. The

resultant 1D-Var analysis with new first guess based on the ANN method showed that the errors of temperature and moisture in the mid-troposphere are reduced, due to the use of larger volume of cloud-affected infrared radiances. Furthermore, the computational time can be substantially reduced as much as 1.85% by the ANN method, compared to the MR method. The evaluation of the ANN method in the UM global weather forecasting system demonstrated that it helps to use more infrared radiances in the cloudy-sky data assimilation. Although its impact on the UM global temperature and moisture forecasts was found to be near neutral, it has been demonstrated that the UM global precipitation forecasts and tropical cyclone forecast, which occur mostly around cloud regions, can be improved by the ANN method.

Keywords: Hyperspectral infrared radiance, IASI, data assimilation, cloudy-sky 1D-Var analysis, artificial neutral network method, NWP forecast

Student number: 2012-20346

Table of Contents

Abstract i

Table of Contents iii

List of Tables v

List of Figures vi

1. Introduction 1

2. Background and theory 6

2.1. IASI hyperspectral measurement 6

2.2. Theoretical background 10

2.3. Radiance simulation in Radiative Transfer Model 12

3. IASI 1D-Var assimilation 14

3.1. Retrieval of cloud top pressure and cloud fraction 14

3.2. 1D-Var analysis 20

4. Preparation of simulation dataset 21

4.1. ECMWF short-range forecast 21

4.2. Simulation of IASI radiances 25

4.3. Simulation of UM background profiles 32

5. Assessment of pre-developed methods with simulation dataset 35

5.1. Pre-developed cloudy-sky radiance assimilation 35

5.2. Assessment of the pre-developed assimilation method 37

6. Development of a new cloud parameters retrieval method 40

6.1. Definition of 'Optimum CTP' 40

6.2. Evaluation of original retrieval method 48

6.3. New retrieval method with an ANN approach 54

7. Assessment of ANN retrieval method in the 1D-Var analysis 58

7.1. Simulation Framework 58

7.2. Experiments with the UM NWP system 73

8. Impact study of ANN method on the UM forecast 83

8.1. Assessment of experiments in the UM NWP system 85

8.2. Impact on the precipitation forecast 92

8.3. Impact of tropical cyclone forecast 97

9. Summary and discussion 110

References 116

국문초록 121

감사의 글 124

List of Tables

Table 1. List of 10 IASI channels used in the MR method.

Table 2. Pressure levels used in the RTTOV model.

Table 3. Number (percentage) of converged and non-converged cases from three methods in the 1D-Var analysis using the simulation dataset.

Table 4. Number (percentage) of converged and non-converged cases from the MR and ANN methods in the 1D-Var analysis of the UM OPS (00 UTC 30 July 2017).

Table 5. Number (percentage) of converged and non-converged cases from the MR and ANN methods in the 1D-Var analysis of the UM OPS, only for the cases having retrieved CTP smaller than 200hPa (00 UTC 30 July 2017).

List of Figures

Figure 1. Collocation of IASI (yellow) and AMSU (red). The distance between adjacent pixels is given for one scan line in km.

Figure 2. Flow chart of IASI 1D-Var assimilation process. Two different cloud retrieval methods, i.e., (a) MR method and (b) ANN method, are compared in this study.

Figure 3. One example showing the residual distribution for all possible pairs of CTP and CF, for the given set of 10 IASI channels. ‘x’s denote CTP and CF pairs obtained by applying Eq. (6) at 27 levels, and ‘Δ’ represents a first guess pair of CTP and CF from the MR method.

Figure 4. (a) Location of selected profiles in the cloud condensate-sampled subsets of the ECMWF short-range forecasts and (b) distribution of the profiles within calendar months. Different subsets are shown in different colors.

Figure 5. Example profiles of (a) temperature, (b) moisture, (c) cloud fraction, (d) cloud liquid water, and (e) ice water content from the ECMWF short-range forecasts.

Figure 6. Standard deviation of simulated IASI observation error using IASI measurement-error covariance matrix (**R**).

Figure 7. Mean of simulated IASI TBs divided by (a) cloud top pressure and (b) cloud fraction. The numbers in parentheses are the number of cases belonging to each group.

Figure 8. Mean of simulated IASI TBs divided by (a) CF >= 0.5 and (b) CF=1. The numbers in parentheses are the number of cases belonging to each group.

Figure 9. Mean of simulated IASI TBs divided by (a) CTP < 500 hPa and (b) CTP >= 500 hPa. The numbers in parentheses are the number of cases belonging to each group.

Figure 10. Mean (black solid line) and standard deviation (red dashed line) profiles of simulated temperature (left column) and moisture (right column) error using UM forecast-error covariance matrix (**B**) at (a) 30° N−90° N, (b) 30° S−30° N, and (c) 30° S− 90° S.

Figure 11. Mean bias (solid line) and RMSE (dashed line) of (a) temperature and (b) humidity profiles analyzed by the cloudy 1D-Var method (black) and the clear channel selection method (blue). Orange line with asterisks represents RMSE of background states.

Figure 12. RMSE' at 27 pairs of CTP and CF denoted by ‘x’s in Figure 2. ‘Δ’ denotes a first guess pair of CTP and CF from the MR method whereas ‘*’ indicates the optimum CTP and CF pair showing the smallest RMSE'.

Figure 13. Schematic diagram showing the method to obtain the optimum cloud parameters. Red dashed box shows the process generating (a) 27 pairs of CTP and CF (as noted ‘x’ marks in Figure 2). (b) RMSE's at 27 RTTOV levels are calculated by comparing the analysis of temperature and moisture to the truth x_t, as exemplified in Figure 3, and the optimum CTP and CF values are determined. The upper index *n* is the ordinal number of total N samples and the lower index *j* represents the RTTOV pressure level. Other notations are found in the text.

Figure 14. Mean bias (solid lines) and RMSE (dashed lines) of temperature (left panels) and humidity (right panels) profiles from the 1D-Var analysis with the use of the optimized CTP and CF from the MR method (blue) and the optimum CTP and CF (red) at (top) 90°N−30°N, (middle) 30°N−30°S, and (bottom) 30°S−90°S. Black dotted lines in the RMSE profiles represent the RMSE profiles of the background state.

Figure 15. Same as Figure 14 except for all latitude regions.

Figure 16. Two-dimensional histograms of frequencies of optimum CTP vs. optimized CTP from the MR method (left panels). Color represents data count. Mean CF differences between optimum and the MR method (right panels) are given in the same CTP pressure coordinates as in the left panels. 90°N−30°N, 30°N−30°S, and 30°S−90°S regions are given from the top to the bottom.

Figure 17. (left) Number of used measurements at each IASI channel from the MR method (blue) and from the optimum CTP and CF (red). (right) Difference in the number of used measurements (optimum minus MR method). Top, middle, and bottom figures represent latitudinal regions of 90°N−30°N, 30°N−30°S, and 30°N−90°S, respectively. Black arrows at bottom figures represent IASI channels in CO_2, window, and water vapor absorption bands.

Figure 18. Mean O-B (observed minus background brightness temperature) for the MR method (blue) and the optimum CTP and CF (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands.

Figure 19. Structure of the ANN model showing 254 inputs (182 IASI TBs, 43-level background temperature and humidity profiles, surface temperature, skin temperature, and surface humidity), a single hidden layer with 5 neurons, and the target output (CTP).

Figure 20. (left) Two-dimensional histograms of frequencies of (a) optimum CTP vs. optimized CTP from the MR method (MR CTP), and (c) optimum CTP vs. optimized CTP from the ANN method (ANN CTP). Color represents data count. (right) Mean CF difference (b) between optimum and the MR method, and (d) between optimum and the ANN method. The mean differences are given in the same CTP pressure coordinates as in the left panels.

Figure 21. (a) Difference in the number of used measurements (ANN method minus MR method) at each of the 182 IASI channels. (b) Mean O-B for the optimum (green), the MR method (blue), and ANN method (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands.

Figure 22. Mean bias (solid lines) and RMSE (dashed lines) of (a) temperature and (b) humidity analysis profiles from the 1D-Var analysis with the use of the MR method (blue), and ANN method (red), and the optimum values (green). Black dotted lines in the RMSE profiles represent the RMSE profiles of the background state.

Figure 23. (Left) Two-dimensional histograms of frequencies of (a) optimum CTP vs. optimized CTP from the MR method (MR CTP), and (c) optimum CTP vs. optimized CTP from the ANN method (ANN CTP). Color represents data count. (right) Mean CF difference (b) between optimum and the MR method, and (d) between optimum and the ANN method. The mean differences are given in the same CTP pressure coordinates as in the left panels. The results for the ANN model in (1) 90°N−30°N, (2) 30°N−30°S, and (3) 30°S−90°S regions are given from the top to the bottom.

Figure 24. (a) Difference in the number of used measurements (ANN method

minus MR method) at each of the 182 IASI channels. (b) Mean O-B for the optimum (green), the MR method (blue), and ANN method (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands. The results for the ANN model in (1) 90°N−30°N, (2) 30°N−30°S, and (3) 30°S−90°S regions are given from the top to the bottom.

Figure 25. Mean bias (solid lines) and RMSE (dashed lines) of (a) temperature and (b) humidity analysis profiles from the 1D-Var analysis with the use of the MR method (blue), and ANN method (red), and the optimum values (green). Black dotted lines in the RMSE profiles represent the RMSE profiles of the background state. The results for the ANN model in (1) 90°N−30°N, (2) 30°N−30°S, and (3) 30°S−90°S regions are given from the top to the bottom.

Figure 26. (a) Two-dimensional histograms of frequencies of ANN CTP vs. MR CTP in the UM OPS system. Color represents data count. (b) Mean difference between MR CF and ANN CF in the UM OPS system. The mean differences are given in the same CTP pressure coordinates as in the left panels.

Figure 27. (a) Difference in the number of used measurements (ANN method minus MR method) at each of the 182 IASI channels. (b) Mean O-B for the optimum (green), MR method (blue), and for ANN method (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands.

Figure 28. RMSE profiles of (a) temperature and (b) relative humidity in the UM OPS analysis with the MR method (blue) and ANN method (red).

Figure 29. RMSE for temperature profiles at T+0 in the control (red) and the experiment (blue) runs at (a) 90°N−30°N, (b) 30°N−30°S, and (c) 30°S−90°S.

Figure 30. RMSE for relative humidity profiles at T+0 in the control (red) and the experiment (blue) runs at (a) 90°N−30°N, (b) 30°N−30°S, and (c) 30°S−90°S.

Figure 31. (a) Two-dimensional histograms of frequencies of ANN CTP vs. MR CTP in the UM OPS experiment excluding the scenes having cloud top above 200 hPa level. Color represents data count. (b) Mean difference

between MR CF and ANN CF in the same experiment. The mean differences are given in the same CTP pressure coordinates as in the left panels.

Figure 32. (a) Difference in the number of used measurements (ANN method minus MR method) at each of the 182 IASI channels in the UM OPS experiment excluding the scenes having cloud top 200 hPa level. (b) Mean O-B for the optimum (green), MR method (blue), and for ANN method (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands.

Figure 33. Mean bias (solid lines) and RMSE (dashed lines) profiles of temperature forecasts at (a) T+0, (b) T+12, (c) T+24, (d) T+36, (e) T+48, (f) T+60, and (g) T+72 in the control (black) and the experiment (red) runs.

Figure 34. Mean bias (solid lines) and RMSE (dashed lines) profiles of humidity forecasts at (a) T+0, (b) T+12, (c) T+24, (d) T+36, (e) T+48, (f) T+60, and (g) T+72 in the control (black) and the experiment (red) runs.

Figure 35. (a) Equitable threat score, (b) false alarm rate, and (c) bias score of global precipitation forecast using GPM IMERG precipitation data as a reference. Blue and red lines indicate the control and experiment runs, respectively.

Figure 36. (a) COMS 10.8 ㎛ image and (b) difference in the number of used IASI channels between in the control run and in the experiment run (experiment minus control) at 00 UTC 01 August. (c) and (d) are at 00 UTC 02 August. Grey line represents the JTWC best track and the black square shows the center of the typhoon Noru.

Figure 37. Typhoon Noru track of the JTWC best track (black with asterisk), the ECMWF reanalysis (green), the KMA forecast (purple), and forecasts from (a) 00 UTC 02 August, (b) 00 UTC 03 August, (c) 00 UTC 04 August, and (d) 00 UTC 05 August 2017 in the control (blue) and the experiment (red) runs. Black dots show every 12-hour track of the whole period of typhoon Noru and triangle and square represent the track at T+0 and T+72 forecast, respectively.

Figure 38. RMSE of (a) minimum sea level pressure and (b) maximum surface wind speed at T+0 to T+72 forecasts for the typhoon Noru in the control

(blue) and experiment (red) runs using the JTWC best track data as a reference. (c) RMSE of precipitation forecasts in both runs is calculated against the GPM IMERG data. Blue (red) represents the results in the control (experiment) run.

Figure 39. Difference between the ERA5 reanalysis and the 36-hour forecast of (a), (b) 850 hPa temperature, (c), (d) 850 hPa humidity, and (e), (f) 850 hPa geopotential height from 00 UTC 04 August in the control (left) and the experiment (right) runs, respectively. The difference is the 36-hour forecast minus ERA5 reanalysis and black line represents the JTWC best track.

Figure 40. (a) GOES-13 10.7 ㎛ image and (b) difference in the number of used IASI channels between in the control run and in the experiment run (experiment minus control) at 00 UTC 07 August. Grey line represents the JTWC best track and the black square shows the center of the hurricane Franklin.

Figure 41. Hurricane Franklin track of the best track (black with asterisk), the ECMWF reanalysis (green), and 72h forecasts from (a) 12 UTC 06 August, (b) 00 UTC 07 August 2017 in the control (blue) and the experiment (red) runs. Black dots show every 12-hour track of the whole period of hurricane Franklin and triangle and square represent the best track at T+0 and T+72 forecast, respectively.

Figure 42. RMSE of (a) minimum sea level pressure and (b) maximum surface wind speed forecast at T+0 to T+72 for the hurricane Franklin in the control (blue) and experiment (red) runs using JTWC best track data as a reference. (c) RMSE of precipitation forecasts in both runs is calculated against the GPM IMERG data. Blue (red) represents the results in the control (experiment) run.

Figure 43. (a) Difference between the ERA5 reanalysis and the 48-hour forecast of 850 hPa geopotential height from 00 UTC 07 August in the control. (b) The difference between the ERA5 reanalysis and the 48-hour forecast of 850 hPa geopotential height from 00 UTC 07 August in the experiment run minus the difference in the control run. The difference means the 48-hour forecast minus ERA5 reanalysis and black line represents the JTWC best track.

1. Introduction

Satellite observations have become indispensable in the current numerical weather prediction (NWP) systems (Bauer et al., 2015; McNally et al., 2014). Amongst satellite observations, hyperspectral infrared sounders such as the Infrared Atmospheric Sounding Interferometer (IASI), the Atmospheric Infrared Sounder (AIRS), and the Cross-track Infrared Sounder (CrIS) have been found to be the single largest contributor to forecasting skill (Hilton et al., 2009; Joo et al., 2013; Le Marshall et al., 2006; McNally et al., 2006; Smith et al., 2009). For this reason, many satellite centers and NWP centers are preparing geostationary hyperspectral infrared sounders such as Infrared Sounder (IRS) on the geosynchronous Meteosat Third Generation (MTG) and Geostationary Interferometric Infrared Sounder (GIIRS) on FY-4A. However, this contribution is mostly due to clear-sky radiance assimilation. Because of the difficulty to correctly simulate upwelling infrared radiances for cloudy conditions, cloudy-sky data assimilation using the hyperspectral infrared sounders is very limited in terms of data volume used and area applied. This is not only due to difficulty in predicting cloud condition itself but also the fact that infrared radiation is easily contaminated by clouds, and the signals are trapped in clouds before they reach the observing sensor. Considering that approximately 60% of the globe is cloud-covered at a given moment (Wylie and Menzel, 1999) and cloudy areas are where

meteorological disturbances actually take place, additional use of hyperspectral radiance data over the cloudy areas in the data assimilation must have a potential to improve forecasting skill. Despite its necessity, there seem to be few studies of actively using hyperspectral infrared measurements under the cloudy condition for data assimilation.

Li et al. (2005) developed a ‘cloud-clearing’ method for the use of AIRS cloud-affected radiance data. This method constructs the clear-sky radiances for the cloudy area that would be measured if it were clear, by modifying cloud-affected hyperspectral infrared radiances with the aid of collocated high-resolution clear-sky information from the Moderate Resolution Imaging Spectroradiometer (MODIS). The resulting cloud-cleared hyperspectral radiances are assimilated by applying a clear-sky assimilation method. However, since clear-sky information is extended to the cloudy area, the constructed hyperspectral infrared radiances tend to be biased toward the clear-sky field.

In the scheme introduced by McNally and Watts (2003), hyperspectral infrared radiances are used over cloudy areas, but only for channels not contaminated by clouds. In this case, channels are deemed to be clear if observed minus simulated channel radiances are smaller than a preset criterion, and then those selected clear channels are used for the assimilation as for clear-sky infrared radiances.

Instead of choosing clear channels (or excluding cloud-contaminated channels), explicit analysis of cloud properties was introduced for cloudy-sky data assimilation by Pavelin et al. (2008). In this technique, for a given field of view, cloud parameters (i.e. cloud top pressure and cloud fraction) are determined by minimizing the difference between observed and simulated infrared radiances under the single-layer grey cloud assumption, before refining those cloud parameter estimates in a one-dimensional variational (1D-Var) analysis. Those obtained cloud parameters are then used for explicitly calculating IASI channel radiances over cloudy areas, for the subsequent four-dimensional variational (4D-Var) assimilation. Although this method allows increased use of partially cloud-affected scenes, it tends to under-estimate the cloud top pressure, leading to a more conservative use of cloud-affected radiances.

More recently, so-called ‘all-sky’ assimilation methods have been developed to use the full cloud-affected hyperspectral radiances regardless of clear-sky or cloudy-sky conditions. Okamoto et al. (2014) takes account of cloud amounts to quantify cloud effect on all-sky hyperspectral infrared radiances to predict each channel’s standard deviation of first guess departure. Then they used the predicted standard deviation to quality control for threshold relaxation, and to observation error estimation for obtaining information from cloud-affected radiances. Moreover, research for all-sky channel selection were carried out to apply them in

both clear-sky and cloudy-sky areas. Considering channels sensitive to cloud microphysical variables, Martinet et al. (2014) added a new all-sky channel set to the pre-existing channel set. Using those all-sky channel set, cloud variables such as cloud liquid/ice water content were even treated as control variables in the all-sky data assimilation. In addition to a channel set used for the clear-sky data assimilation, a humidity-sensitive-channel set was added to assimilate hyperspectral infrared radiances in all-sky conditions (Migliorini, 2015). However, success has been limited mainly because current NWP and radiative transfer models are not realistic enough to represent accurately the three-dimensional nature of clouds and their radiative influences, which are essential for successful implementation of the all-sky assimilation method.

In this study, aimed at the improvement of 1D-Var analyses over cloudy area, we attempt to improve the Cloudy 1D-Var method developed by Pavelin et al. (2008). It is difficult to evaluate the retrieved cloud properties based on the assumption of opaque and single-layer cloud because such assumed cloud properties are mostly far from the observed cloud properties. Furthermore, although the retrieved cloud parameters in the Cloudy 1D-Var method play an important role not only in simulating radiances but also in selecting channels used for the following 1D-Var analysis, it has not been examined whether these retrieved cloud parameters are the best solutions for the 1D-Var analysis. In other

words, there might be different solutions that lead to better 1D-Var analysis even if the single-layer grey-body cloud assumption is made.

This study attempts to improve 1D-Var analysis over cloudy areas by introducing the best solutions that yield the optimal 1D-Var analysis of temperature and humidity under the single-layer grey-body cloud assumption. First, we find the ‘optimum’ cloud top pressure and cloud fraction that give the best temperature and humidity results in the 1D-Var analysis. Then, we devise a way to retrieve cloud parameters as close as possible to those optimum solutions from IASI brightness temperatures (TBs) plus model background information. The newly developed retrieval method will be tested with an European Centre for Medium-Range Weather Forecasts (ECMWF) short-range forecast based simulation dataset to evaluate whether it provides a positive impact on the 1D-Var analysis. Furthermore, the impact of the new method will be examined for the possibility of implementation in the operational preprocessor of the Unified Model (UM) NWP system.

2. Background and theory

2.1. IASI hyperspectral measurement

IASI is one of the hyperspectral infrared sounding instruments and it is carried on board the European meteorological polar-orbiting satellites, Metop-A (launched on 19 October 2006), Metop-B (launched on 17 September 2012), and Metop-C (launched on 7 November 2018). The main purpose of the IASI mission is to support NWP model by providing upwelling spectra to produce temperature and humidity profiles with high vertical resolution and accuracy (Klaes et al., 2007; Lerner et al., 2002). In fact, IASI radiances have been utilized for a number of years making a great contribution to many NWP forecast systems (August et al., 2012; Prunet et al., 1998). IASI is also used to retrieve the hyperspectral resolution soundings of ozone, nitrous oxide, carbon dioxide, and methane, and to monitor the changes of the gas profiles.

The effective field of view (EFOV) contains 2 x 2 circular pixel matrix of instantaneous field of view (IFOV) with about 12 km diameter footprint at nadir (Figure 1). The cross-scan width of the instrument is about 2,200 km with the range of 48° 20′ on either side of the nadir direction. Providing global observation twice a day with 14 orbits, IASI has sun-synchronous mid-morning orbit with descending node at 9:30 am local solar time. IASI is a Fourier interferometer with

8461 channels in the infrared part of the electromagnetic spectrum from 645 cm^{-1} and 2760 cm^{-1} with a spectral resolution 0.25 cm^{-1} after apodization. The noise equivalent temperature difference of the raw spectrum is less than 0.3 K for wavenumbers below 2200 cm^{-1} and the difference is larger in the wavenumbers above 2500 cm^{-1}.

The IASI spectral range was chosen to include the CO_2, water vapor, ozone and other trace gases absorption bands and some solar channels. The window region is also included to derive the surface temperature and emissivity over land and sea, and some cloud properties. Additionally, the interaction between atmospheric states and the clouds can be derived with its window channels. From IASI thermal infrared spectra, temperature and humidity profiles with high vertical resolution can be retrieved with an accuracy of absolute 1 K and relative 10%, respectively (Diebel et al., 1996).

In the UM Observation Processing System (OPS), processing the IASI radiance data starts from the collocation of Advanced Microwave Sounding Unit (AMSU) data to the IASI footprints. Firstly, surface type is assigned to use the collocated AMSU data by matching the TBs against eight known surface types. Then cloud detection method is applied using not only the IASI TBs but also the collocated AMSU TBs in three different ways. Either if the standard deviation of the IASI TBs at 2390 cm^{-1} in one footprint (four pixels) is greater than a threshold

or if the TB difference between predicted TB at 2390 cm^{-1} is larger than a pre-defined threshold, the pixel is considered as cloudy, or AMSU cloud detection method can be adapted in this step. When the scene is classed as clear, the IASI radiances in 314 channels (Collard, 2007) for the most likely homogeneous and clear pixel among the four pixels are passed and used for the data assimilation process in the UM OPS.

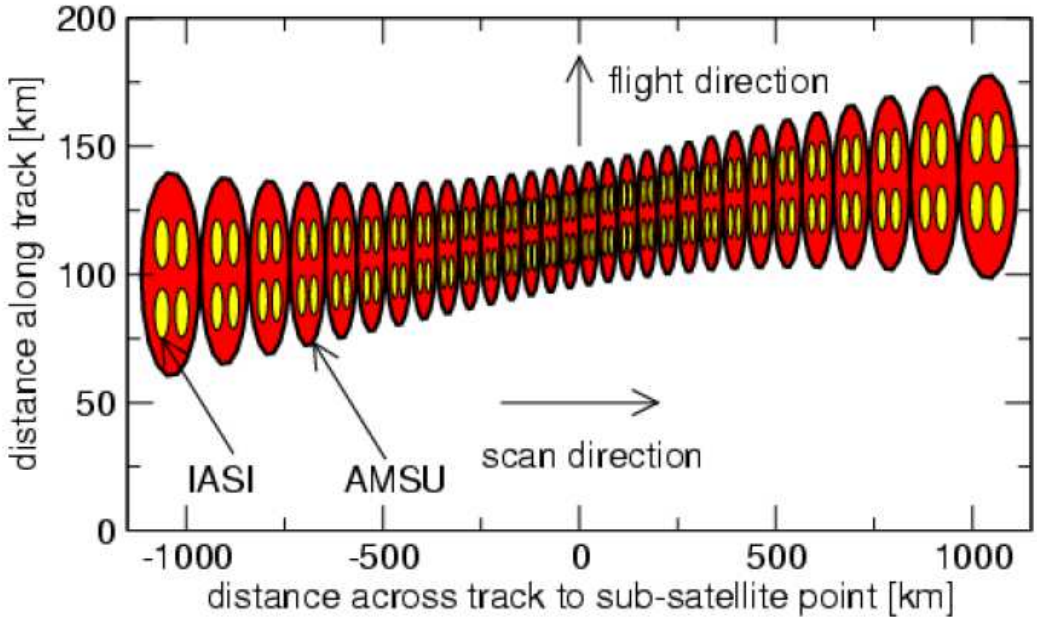


Figure 1. Collocation of IASI (yellow) and AMSU (red). The distance between adjacent pixels is given for one scan line in km.

2.2. Theoretical background

The 1D-Var analysis based on the Bayesian optimal estimation is used in the UM OPS before IASI radiances are passed to the 4D-Var process for data assimilation. In this step, when the atmospheric soundings are not converged, the IASI channel radiances are considered to be unacceptable and screened out. Basically, the 1D-Var technique helps to find a statistically optimal estimate of the atmospheric states, x, by minimizing the cost function described by Rogers (2000):

$$J(x) = (x - x_0)^T B^{-1}(x - x_0) + \{y - y(x)\}^T R^{-1}\{y - y(x)\} \quad (1)$$

where y is the observation, x_0 is the background state, **R** is the observation error covariance matrix, and **B** is the background error covariance matrix. Here, $y(x)$ is the forward-modelled radiances from the atmospheric states, x, and the **R** matrix includes the forward model errors. When there are larger number of channels than the elements of atmospheric states, the iterative solution is

$$x_{n+1} = x_0 + (\mathbf{B^{-1}} + \mathbf{H_n}^{\mathbf{T}}\mathbf{R^{-1}}\mathbf{H_n})^{-1}\mathbf{H_n}^{\mathbf{T}}\mathbf{R^{-1}}[y - y(x_n) - \mathbf{H_n}(x_0 - x_n)] \quad (2)$$

where x_n is nth estimate of the atmospheric states and $\mathbf{H_n} = \nabla_x y(x_n)$. Either Newtonian minimization with the certain step size or Marquardt-Levenberg

descent algorithm with the best step size for each iteration can be used in the minimization process. In the UM OPS, the forward-modelled IASI radiances are simulated by the Radiative Transfer for TIROS Operational Vertical Sounder (RTTOV) model.

It is essential to consider additional cloud variables in the atmospheric state vector to assimilate all-sky IASI radiances. Assuming that all atmospheric conditions can be interpreted with a simple cloud, in the UM 1D-Var system, cloud top pressure (CTP) and cloud fraction (CF) are added to the atmospheric states and simple cloud scheme in RTTOV model is used. Here, 'grey-body cloud assumption' is applied, that is defined as cloud emissivity has no dependency on wavelength as follows:

$$\frac{\delta \varepsilon}{\delta \lambda} = 0 \tag{3}$$

where ε is the emissivity and λ is the wavelength. Therefore, cloud effect of all channels is only proportional to geometric fraction of the cloud, not to optical property depending on the wavelength within infrared range.

2.3. Radiance simulation in Radiative Transfer Model

RTTOV is a fast regression model calculating radiance of infrared and microwave sounders from atmospheric profiles such as temperature, water vapor, ozone, carbon dioxide, methane, cloud and surface properties. This radiative transfer model calculates changes of the top of the atmosphere (TOA) radiance for each instrument channels given unit perturbation of the atmospheric states, which is called as Jacobian. The Jacobian is obtained for each profile as the perturbation is based on the input atmospheric profiles, assuming that relationship between the radiance changes and the variable perturbations is likely linear. In the variational assimilation, the RTTOV model provides changes in the atmospheric states given TOA radiance perturbations.

Cloudy-affected infrared radiance is simulated in the RTTOV model by using two different schemes; simple cloud scheme and multiple scattering scheme. In the simple cloud scheme, cloud is assumed to be optically thick (opaque) single-layer water cloud with a given cloud top pressure and an 'effective' cloud fraction so that the cloudy-sky radiance R_{cloudy} is calculated as follow:

$$R_{cloudy} = (1 - N_e)R_{clear} + N_e R_{overcast}(p_c) \quad (4)$$

where R_{clear} is the clear-sky radiance, $R_{overcast}$ is the radiance under overcast

condition with cloud top at pressure P_c, and N_e is the effective cloud fraction. The effective cloud fraction means the inner product of the geometrical cloud fraction and the cloud emissivity (i.e. the unity) with the single-layer grey-body cloud assumption.

Secondly, more complex cloud types and multi-layer clouds are considered in the multi scattering scheme. In this scheme, backscattering and transmission between layers are considered to scale the optical thickness, which is essentially used for calculating transmittance. Five different types of water cloud (Stratus Continental, Stratus Maritime, Cumulus Continental Clean, Cumulus Continental Polluted, and Cumulus Maritime) and 30 types of ice cloud for different shapes of the ice crystal (hexagonal and aggregates) are available. Using the additional inputs of profiles of cloud liquid water and cloud ice water contents, scattering effect is parameterized. For water clouds, total number concentration is derived from liquid water content, and for ice clouds, effective diameter of ice crystal is derived using temperature or ice water content with different options (Ou and Liou (1995), Wyser (1998), Boudala et al (2002), and McFarquar et al (2003)). In order to combine the cloud-free radiance and cloudy radiance in one atmospheric profile, the stream method (Amorati and Rizzi, 2002) is used. The atmosphere is divided into a number of columns and the columns are sharing the cloud fraction of each layer, so that overlapping cloud can be considered at the end.

3. IASI 1D-Var assimilation

IASI hyperspectral radiances have been subjected in the UM OPS before being used in the 4D-Var data assimilation system in the UM (Rawlins et al., 2007). In order to evaluate the potential performance of the scheme in the 4D-Var system, a 1D-Var analysis is used as a simple method in this study. In this section, we briefly introduce the Cloudy 1D-Var method (Pavelin et al, 2008) in the IASI 1D-Var assimilation. The flowchart in Figure 2 summarizes the steps of the IASI 1D-Var assimilation.

3.1. Retrieval of cloud top pressure and cloud fraction

In the Cloudy 1D-Var method, the assumed single-layer grey-body cloud is expressed with CTP and CF. For the calculation of the cloud's radiative effect within the given IASI field of view, a first guess pair of CTP and CF is provided. They are simultaneously retrieved by the minimum residual method (Eyre and Menzel, 1989) (hereafter referred to as MR method), as expressed as the method (a) in the dashed rectangle of Figure 2. The first guess pair is determined from the observed and simulated radiances at 10 IASI channels given in Table 1. These 10 channels were selected to locate the channel's weighting functions evenly throughout the troposphere so that the cloud's influences on all the radiances are represented.

Figure 3 shows an example of the residual between observed and simulated radiances, cumulated over the 10 IASI channels, for all given CTP and CF pairs. For a CTP and CF pair, the residual is defined as follows:

$$Residual = \sum_{i=1}^{10} [R_{obs}^{i} - H(\mathbf{x_b}, p_c, N)]^2 = \sum_{i=1}^{10} \{R_{obs}^{i} - [(1-N)R_{clr}^{i} + NR_{cld}^{i}(p_c)]\}^2 \quad (5)$$

where R_{obs} is the observed radiance; H is a forward operator; $\mathbf{x}_b$ is the background state; p_c and N are the cloud top pressure and cloud fraction, respectively; R_{clr} is the simulated clear-sky radiance; $R_{cld}(p_c)$ is the simulated radiance for the overcast sky whose cloud top pressure is p_c; i is an index for the 10 IASI channels. When the MR method finds one pair of CTP and CF having the smallest residual, for practical purposes to reduce the computational burden, each RTTOV-pressure level (Table 2) is assigned as a CTP and a corresponding CF is solved, i.e.:

$$N = \frac{\sum_{i=1}^{10} (R_{cld}^{i}(p_c) - R_{clr}^{i})(R_{obs}^{i} - R_{clr}^{i})}{\sum_{i=1}^{10} (R_{cld}^{i}(p_c) - R_{clr}^{i})^2} \quad (6)$$

Thus, at each RTTOV pressure level, a CTP and CF pair satisfying the single-layer grey cloud assumption is found (marked by ‘x’ in Figure 3). After finding pairs of CTP and CF at all 27 RTTOV pressure levels, a pair showing the overall minimum residual (marked by ‘Δ’ in Figure 3) is found.

The obtained CTP and CF are then used as initial guess values for optimizing CTP and CF with background atmospheric variables in 1D-Var. This time, 182 operational IASI channels are used for producing optimized CTP and CF as outputs. In this optimization process, the UM operational 6-hour forecast-error covariance matrix (**B**) and the clear-sky observation error covariance matrix (**R**) are used. Diagonal and off-diagonal components of temperature, humidity, and surface states exist in the background error covariance while only diagonal components are available for CTP and CF. Values for diagonal components of background error covariance for CTP and CF are 1000^2 and 1^2, respectively. For the observation error covariance, only the diagonal component is used for 182 channels. If the initial CTP and CF values are not reasonable, the cost function in the 1D-Var will become too large and the solution will not converge. If a IASI pixel results in such a failure, then the scene is considered to be non-converged and the IASI radiances will not be assimilated.

Table 1. List of 10 IASI channels used in the MR method.

Channel number	Wavenumber [cm^{-1}]	Channel number	Wavenumber [cm^{-1}]
179	689.50	280	714.75
205	696.00	327	726.50
212	697.75	407	746.50
242	705.25	662	810.25
246	706.25	1786	1091.25

Table 2. Pressure levels used in the RTTOV model.

Index	Pressure [hPa]	Index	Pressure [hPa]	Index	Pressure [hPa]
1	102.05	10	358.28	19	749.12
2	122.04	11	396.81	20	795.09
3	143.84	12	436.95	21	839.95
4	167.95	13	478.54	22	882.80
5	194.36	14	521.46	23	922.46
6	222.94	15	565.54	24	957.44
7	253.71	16	610.60	25	985.88
8	286.60	17	656.43	26	1005.43
9	321.50	18	702.73	27	1013.25

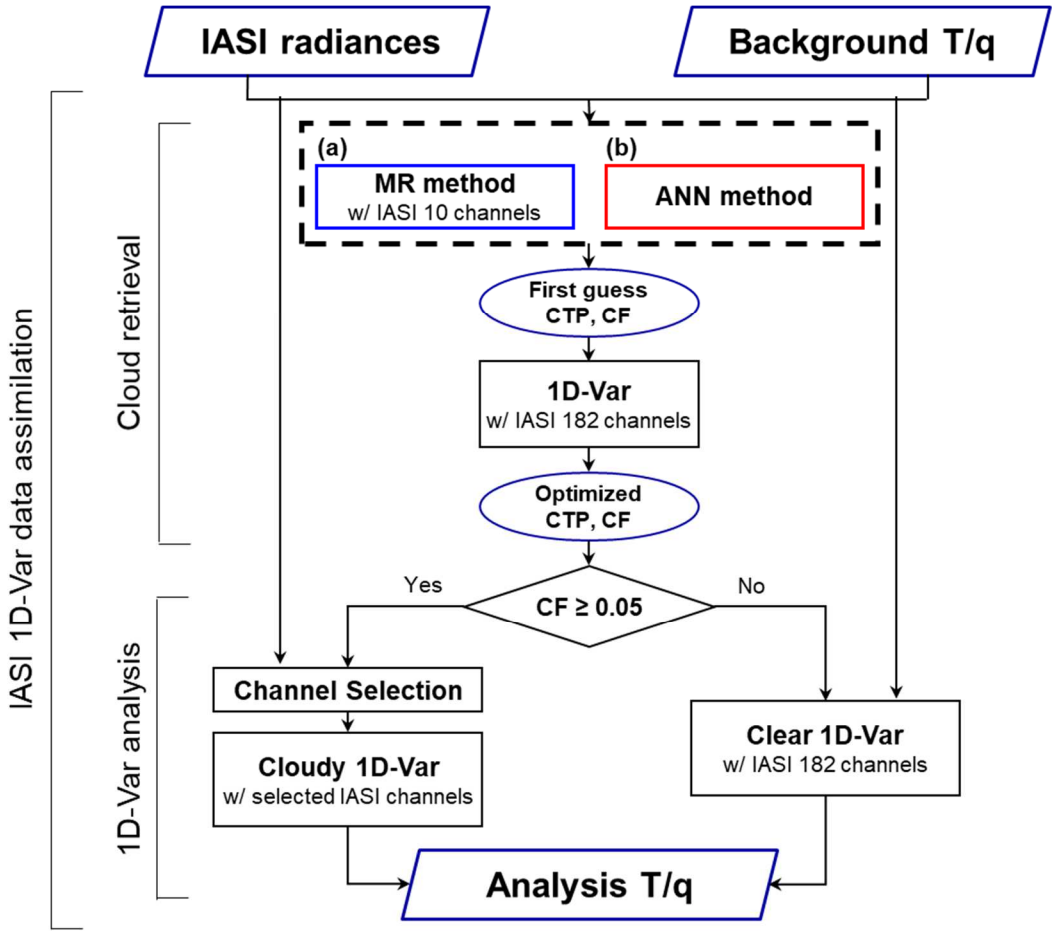


Figure 2. Flow chart of IASI 1D-Var assimilation process. Two different cloud retrieval methods, i.e., (a) MR method and (b) ANN method, are compared in this study.

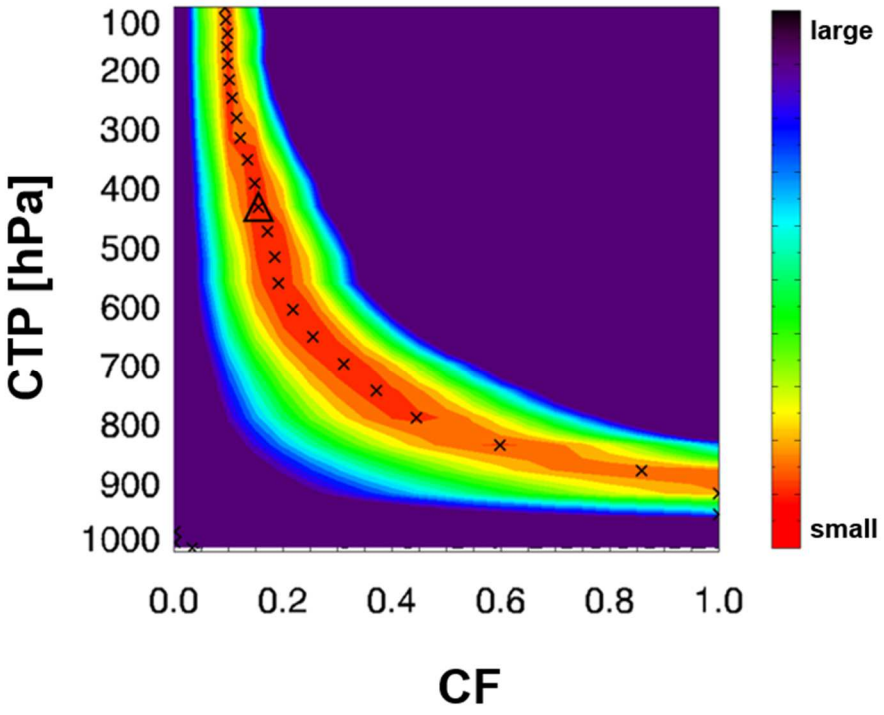


Figure 3. One example showing the residual distribution for all possible pairs of CTP and CF, for the given set of 10 IASI channels. 'x's denote CTP and CF pairs obtained by applying Eq. (6) at 27 levels, and 'Δ' represents a first guess pair of CTP and CF from the MR method.

3.2. 1D-Var analysis

Once the optimized CTP and CF are obtained at each IASI pixel, the assimilation type is determined based on the CF value. If the optimized CF is larger than 0.05, the field of view is considered to be cloudy. Otherwise, it is considered to be clear. If the pixel is determined to be cloudy, channels are selected among the 182 channels for further assimilation if the integrated temperature Jacobian value above the optimized CTP height is greater than 90% of the total Jacobian value. Thus, either 182 channels for clear conditions or selected channels under cloudy conditions are used in the 1D-Var assimilation. In this 1D-Var process, the UM operational **B** matrix and the clear-sky **R** matrix are used. If the case does not converge in the 1D-Var analysis process, the scene is then rejected in the assimilation. The procedures of retrieval of CTP and CF, cloudy (or clear) scene determination, and 1D-Var process given in Figure 2 is referred to as the IASI 1D-Var data assimilation. This assimilation process is applied only over open ocean due to the uncertainty of the surface emissivity over land and sea ice.

4. Preparation of simulation dataset

4.1. ECMWF short-range forecast

Under cloudy-sky condition, it is rather difficult to find observed IASI measurement and collocated real-observed atmospheric profiles. Also, we hardly get the collocated cloud information, for example, cloud fraction and cloud water contents profiles by real observation. So we made a simulation dataset not only to assess the original IASI 1D-Var data assimilation process but also to develop a new retrieval method. Datasets are obtained from 25,000 atmospheric profiles which were compiled from the ECMWF global operational short-range forecasts spanning the time period of 1 September 2013 – 31 August 2014 (Eresmaa and McNally, 2014). These profiles were compiled through a randomized selection to preserve global and seasonal statistical features that the original forecast data hold (Chevallier et al, 2006). Spatial and temporal distributions of the data are homogeneous as shown in Figure 4.

Out of 25,000 profiles, 14,804 profiles were selected for cloudy cases over open ocean, after excluding land profiles. Atmospheric profiles of about a half of cloudy-sky profiles (i.e. 7,498 out of 14,804) were duplicated for constructing the clear-sky profiles (by assuming zero cloud liquid water), after considering the global cloud occurrence of around 70% (Wylie and Menzel, 1999). Thus total

samples are 22,302 profiles. The original ECMWF 14,804 samples over open ocean are not enough because we need cloudy and clear conditions, as well as training and validation datasets. Thus, constructed clear-sky profiles are assumed to be from locations very close to those for cloudy-sky profiles. However, these two datasets may become independent once simulated IASI brightness temperatures are binned with profiles.

The data include temperature and humidity profiles, and vertical distributions of cloud fraction and cloud liquid/ice water content in both northern and southern hemispheres. The examples of the profiles are shown in Figure 5. In this study, these data are considered to be true (referred to as 'truth $\mathbf{x}_t$') and used for simulating IASI-observed radiances and model background states, following the methodology used by Pavelin et al. (2008).

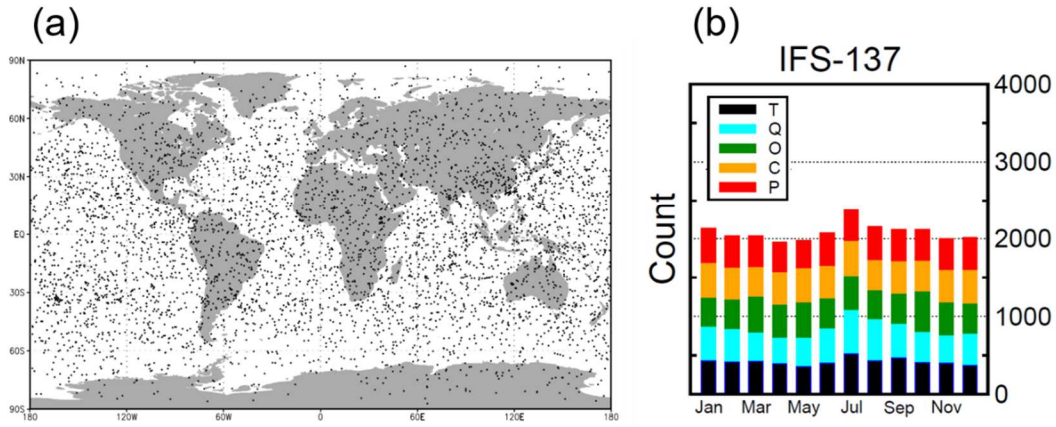


Figure 4. (a) Location of selected profiles in the cloud condensate-sampled subsets of the ECMWF short-range forecasts and (b) distribution of the profiles within calendar months. Different subsets are shown in different colors.

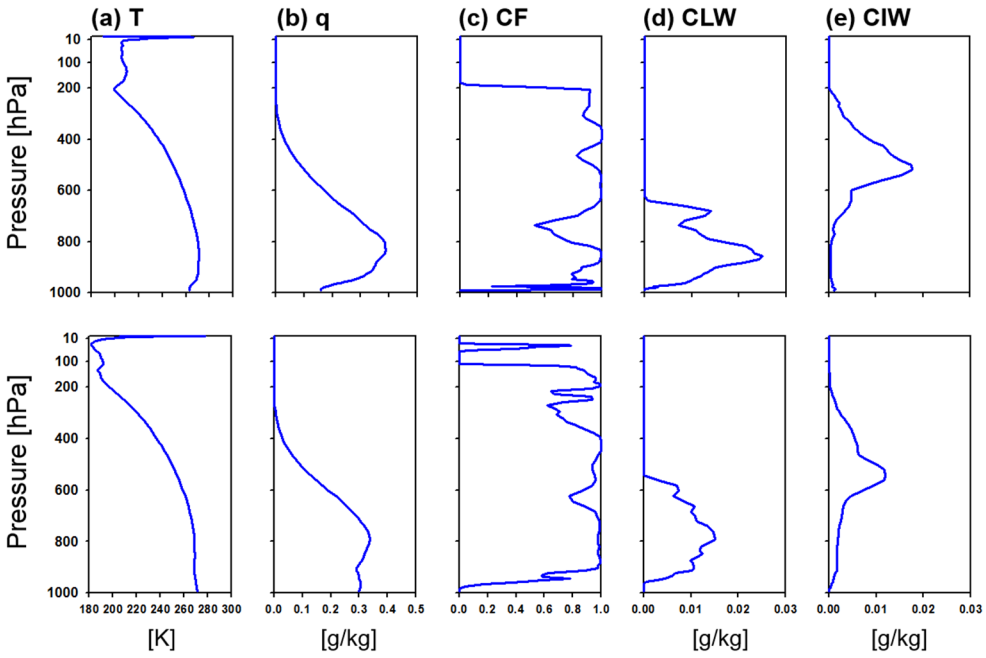


Figure 5. Example profiles of (a) temperature, (b) moisture, (c) cloud fraction, (d) cloud liquid water, and (e) ice water content from the ECMWF short-range forecasts.

4.2. Simulation of IASI radiances

The simulation of IASI radiances at 182 channels (**y**), which are used in the operational UM IASI data assimilation system, is done as follows:

$$\mathbf{y} = H(\mathbf{x}_t) + \mathbf{R} \tag{7}$$

Where **R** is the observation error covariance matrix. The RTTOV version 9.3 model was used as a forward operator *H* in this study. In addition to the truth temperature and humidity profiles used as inputs to the RTTOV model, the cloud profiles from the ECMWF forecasts are taken for the cloudy case simulations. Stratus and cumulus maritime cloud types are chosen for the water cloud simulation. For the ice cloud simulation, McFarquar et al. (2003) scheme and hexagonal ice crystal shape, which is known to be more similar to the real IASI observation, are chosen (Faijan et al., 2012).

After conducting the radiative transfer calculations, expected observation errors are added to simulated IASI brightness temperatures by assuming that observation errors are random with an unbiased Gaussian distribution. Observation errors are obtained from the diagonal component of the clear-sky **R** matrix used in the operational UM OPS. Figure 6 shows the standard deviation of the difference between simulated IASI TBs with the errors and the simulated TBs without the errors.

In order to examine the simulation results of the IASI cloudy TBs, we divided them by different CTP and CF. Here the CTP is defined by the lowest pressure at which the cloud fraction exceeds 10% of its maximum and the CF is defined by the maximum value of its profile (Pavelin et al., 2008). Figure 7 shows the mean of the simulated IASI TBs for (a) the 8 different ranges of CTPs and (b) the 10 different ranges of CFs over the whole IASI wavenumbers. The CTP range of the sample was classified by 150−250, 250−350, 350−450, 450−550, 550−650, 650−750, 750−850, 850−950 hPa and the CF range was 0.001−0.1, 0.1−0.2, 0.2−0.3, 0.3−0.4, 0.4−0.5, 0.5−0.6, 0.6−0.7, 0.7−0.8, 0.8−0.9, 0.9−1.0. The cases having lower CTP have higher mean TB throughout the all wavenumbers and the TB difference between the cases with the lowest CTP and the highest CTP is up to 40 K. It is also demonstrated that small CF makes higher TB because the surface and lower-level atmospheric contribution to the TB gets large. In both figures, the biggest difference is over the range of 800-1000 cm^{-1} while the mean of the TB over the CO_2 and O_3 bands does not change at that much.

We also investigated the sensitivity of TB to CTP when there is a large or small area of cloud. If there is a large amount of clouds, the simulated TB values would be more dependent on the cloud temperature or altitude itself, so that the TB difference according to the altitude of cloud would be noticeable. This was confirmed Figure 8 by the mean of simulated IASI TBs for the cases having CF

>= 0.5 and CF=1. It is shown that the difference of simulated TBs for the cloud top height becomes more pronounced when the cloud amount is large. The simulated results are divided into upper and lower cloud groups to see under what conditions the TB difference by CF is large (Figure 9). The sensitivity test was performed by dividing the high cloud cases (Figure 9(a)) and low cloud cases (Figure 9(b)) using above CTP definition. The higher cloud group has higher sensitivity of TB by the CF value. This explains that the temperature of high cloud has a lower value than surrounding atmospheric temperature compared to the temperature of low cloud, and the change in TB is largely dependent on the amount of clouds. Overall, all of these features in Figure 7, Figure 8, and Figure 9 indicate that the simulated IASI TBs in this study are reasonable to be used.

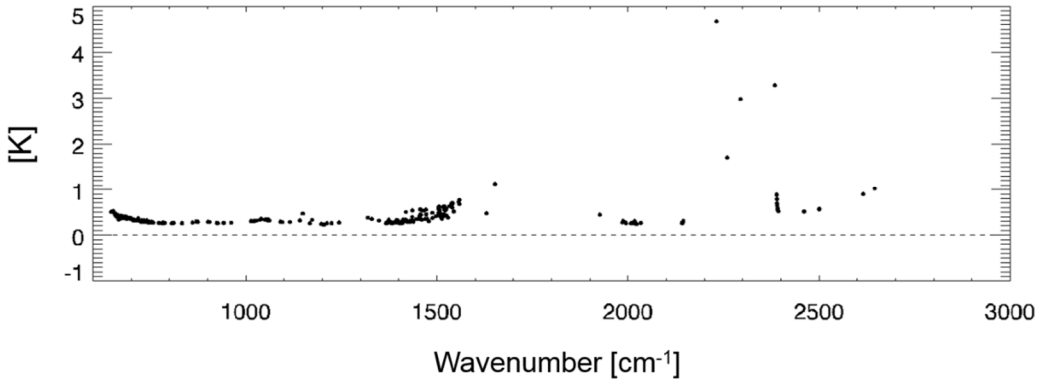


Figure 6. Standard deviation of simulated IASI observation error using IASI measurement-error covariance matrix (**R**).

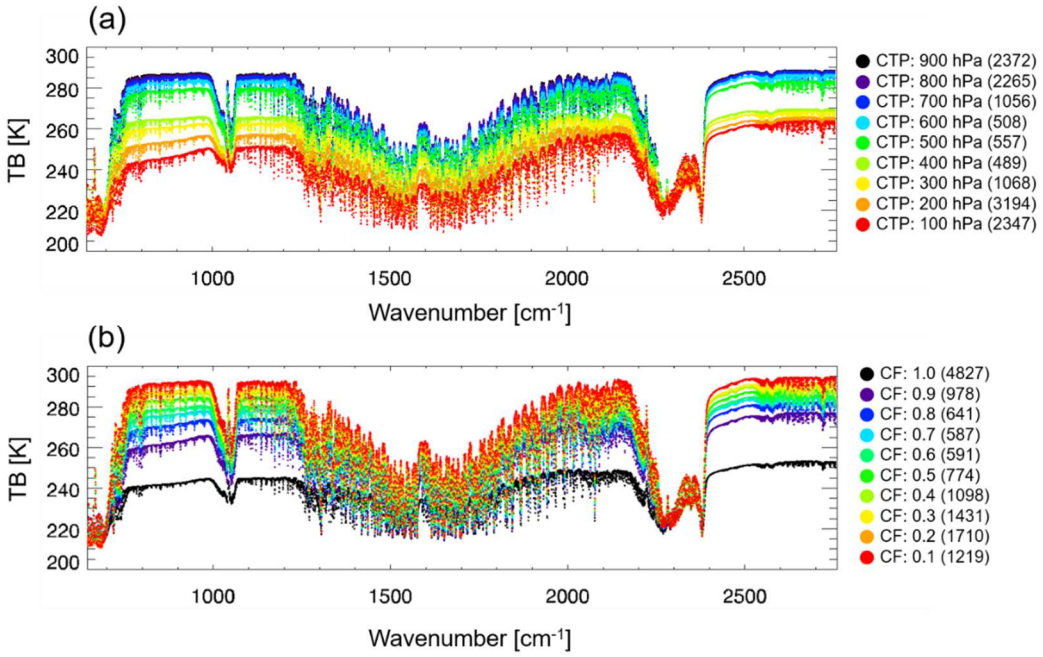


Figure 7. Mean of simulated IASI TBs divided by (a) cloud top pressure and (b) cloud fraction. The numbers in parentheses are the number of cases belonging to each group.

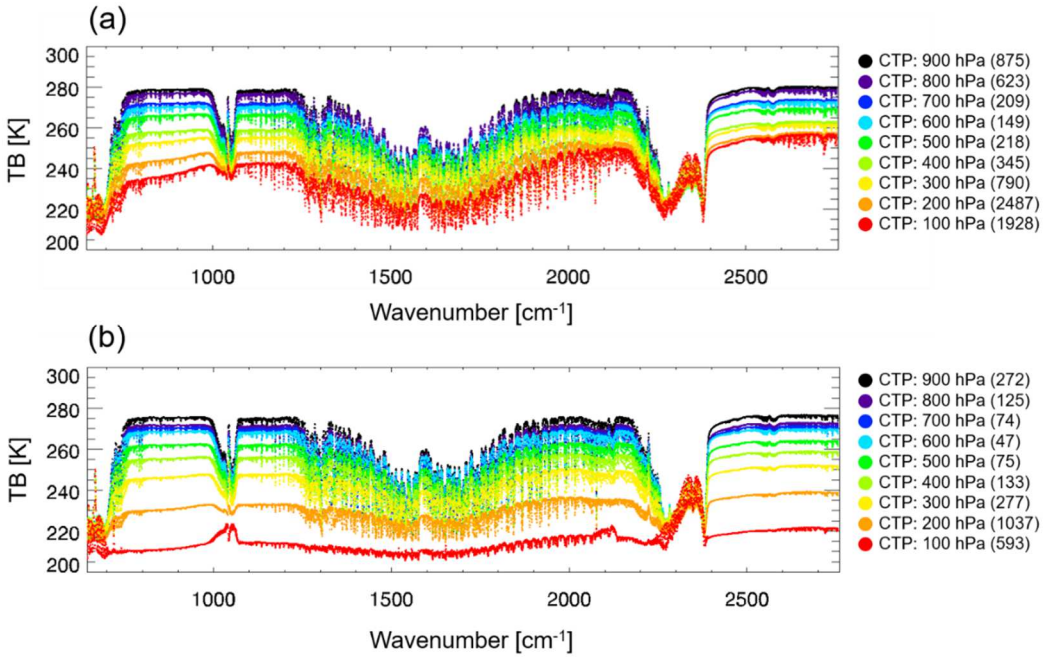


Figure 8. Mean of simulated IASI TBs divided by (a) CF >= 0.5 and (b) CF=1. The numbers in parentheses are the number of cases belonging to each group.

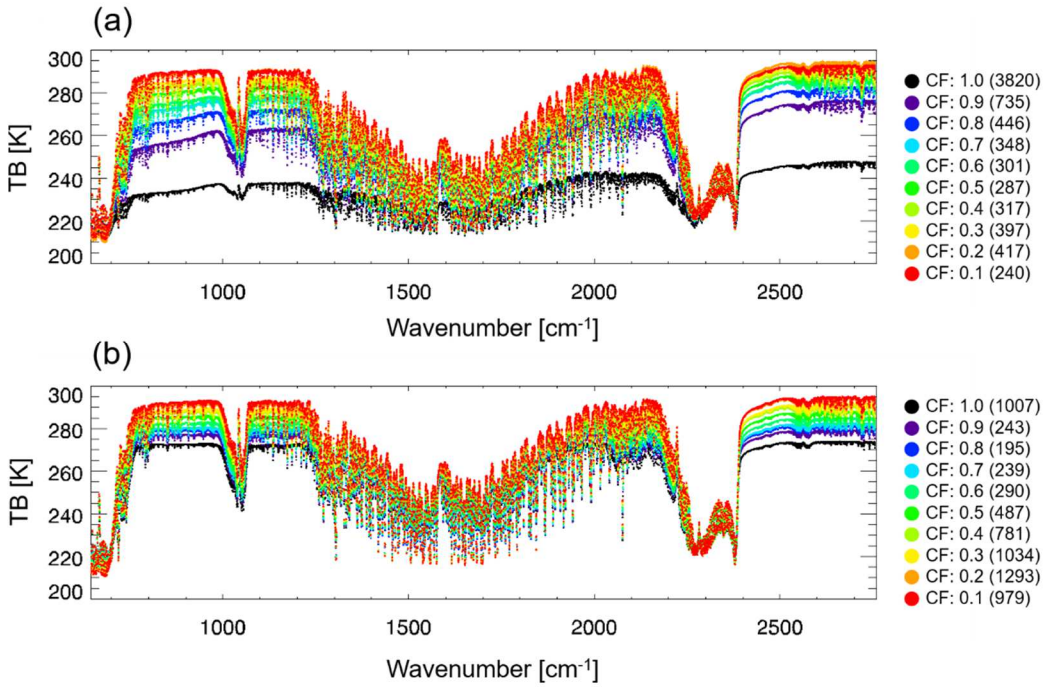


Figure 9. Mean of simulated IASI TBs divided by (a) CTP < 500 hPa and (b) CTP >= 500 hPa. The numbers in parentheses are the number of cases belonging to each group.

4.3. Simulation of UM background profiles

In addition to the IASI radiance simulations, the development procedures leading to temperature and humidity analysis through the 1D-Var process require the model background atmospheric state. Since ECMWF forecasts were treated as the truth, random forecast errors from the UM operational 6-hour forecast-error covariance matrix (**B**) are added to the truth profiles, i.e.:

$$x_0 = x_t + \sum_i \varepsilon_i \lambda_i^{\frac{1}{2}} \mathbf{I}_i \qquad (8)$$

where x_0 is the background profile, x_t is the reference profile, $\mathbf{I}_i$ and λ_i are the eigenvector and eigenvalue of the **B** matrix, and ε_i is a random number having a zero mean and a unit standard deviation (Pavelin et al., 2008). The background fields comprise temperature and humidity profiles, surface temperature, surface humidity, and skin temperature. The **B** matrix is different for latitudinal bands (90°N−30°N, 30°N−30°S, and 30°S−90°S). Figure 10 describes bias and standard deviation of difference of temperature and moisture profiles between simulated background and the true states for each latitude region.

In this study, since the **B** matrices in the UM OPS are only available at three latitudinal regions, trainings of the new retrieval method are taken at the same three latitudinal regions. However, because 22,302 samples are not enough for

training as well as validation for three different regions, the same samples were used for developing the new method at three regions, but with regionally different background errors. It might be thought that extreme profiles at one region (such as profiles reflecting humid condition in 30°N−30°S region) may not fit in conditions for another region (such as 90°N−30°N region). We calculated statistical ranges for the subset of 90°N−30°N and estimated what percentage of total oceanic 14,804 samples is beyond the statistical range of the 90°N−30°N subset. In doing so, 500 hPa level temperature and 850 hPa level humidity were chosen as a reference. From the calculation, it is found that only small percentages of temperature (about 1%, 154 out of total 14,804 samples) and humidity (about 2%, 301 out of total 14,804 samples) are beyond 2σ range of the 90°N−30°N subset ($\mu - 2\sigma \leq X \leq \mu + 2\sigma$, $Pr \approx 95\%$). Near zero percentages of temperature and humidity are beyond 3σ range. Finding that the extreme ranges are quite similar to each other amongst three regions, due to the dataset covering the one-year period including the summer and winter, the current approach of using the same global data for three latitudinal regions is considered to be reasonable. Thus, the following process is equally applied with the same 22,302 profiles for three regions.

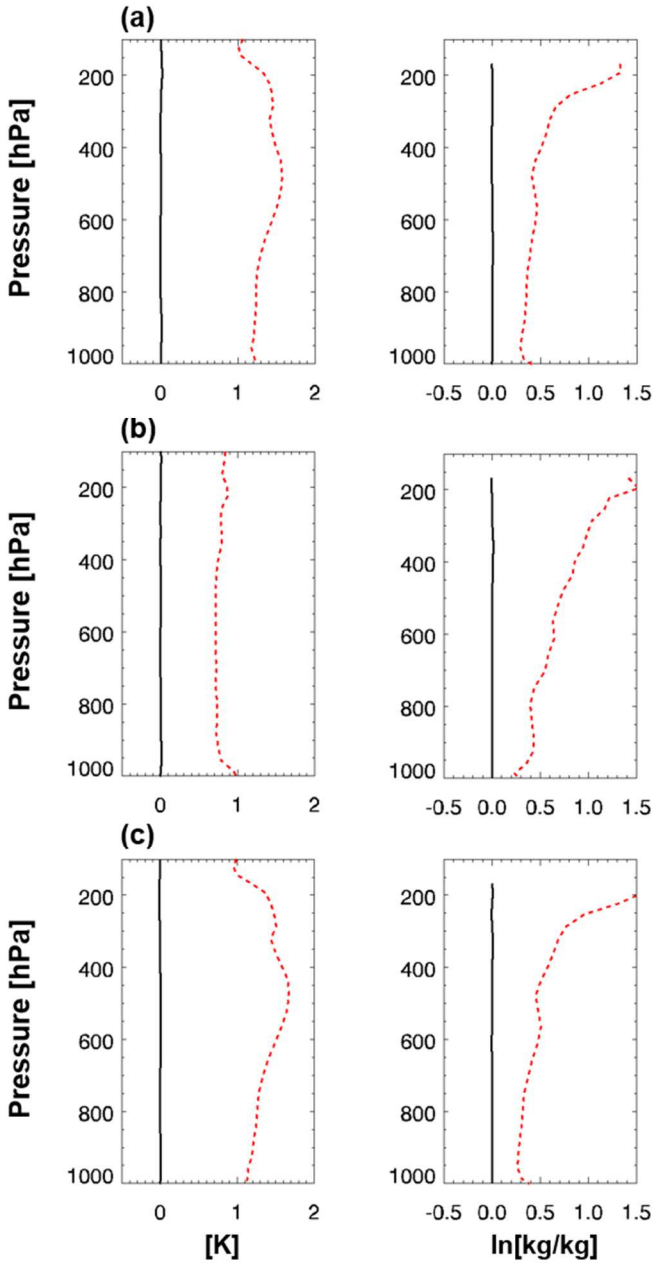


Figure 10. Mean (black solid line) and standard deviation (red dashed line) profiles of simulated temperature (left column) and moisture (right column) error using UM forecast-error covariance matrix (**B**) at (a) 30° N−90° N, (b) 30° S−30° N, and (c) 30° S− 90° S.

5. Assessment of pre-developed methods with simulation dataset

5.1. Pre-developed cloudy-sky radiance assimilation

As it was mentioned, there was an attempt to assimilate some clear channels within a cloudy radiance spectrum (McNally and Watts, 2003). Based on the difference between observed radiance and forward-modelled radiance of each infrared channel, this method detects cloud contamination in each radiance rather than in a complete field of view. The radiance differences are ranked on the basis of the altitude at which each channel begins to be affected by cloud. The channels sensitive to high cloud have smaller channel index and the channels sensitive to low cloud have larger index. The channel ranking is determined within each observed scene. After the TB departures are divided into 5 different bands (long-wave CO_2, water-vapor, O_3 and two short-wave CO_2 bands), a low-pass filter is applied in order to avoid some other noises. Once the cloud signal at a channel is more than a pre-defined threshold, CTP is determined as the level at which the channel is sensitive. Afterward, this detected channel and the channels having larger index are considered as cloud contaminated channels and rejected from the clear-sky radiance data assimilation.

It has been pointed out the limitations of the selection of clear channels over

cloudy condition since most of the infrared channels are rejected within real cloudy scene. Instead of assimilating clear-sky radiances with the limited channels, a direct assimilation of the cloud-affected radiances has been suggested to take advantage of as much channels as possible (Pavelin et al., 2008). As described in Section 3.2, only channels with more than 10% of their whole integrated temperature Jacobian value below CTP are removed in this method. The retrieved cloud parameters such as CTP and CF are assumed to interpret the remaining cloud-affected radiances with the grey-body cloud assumption and cloudy 1D-Var analysis is performed with those values fixed. This algorithm was a starting point of utilizing cloud-affected infrared radiances in direct cloudy-sky data assimilation although it still has limitation on the use in real cloud situations.

5.2. Assessment of the pre-developed assimilation method

In Pavelin et al. (2008), it had been discussed that the Cloudy 1D-Var method performs better in terms of analysis of temperature and moisture profiles than the clear channel selection method (McNally and Watts, 2003). In order to confirm the comparison result in our simulation framework, the analysis of the temperature and humidity profiles from the Cloudy 1D-Var method and clear channel selection method was investigated. Of the total 60,618 simulated profile dataset over sea, 37,029, 16,134 profiles were passed in the cloudy 1D-Var and clear 1D-Var analysis, respectively. The rejected soundings from the 1D-Var assimilation are due to the non-convergence and the analysis result from those cases they are replaced with the corresponding background profiles. As the clear channel selection method is performed for the clear-sky 1D-Var analysis, 52,448 clear-sky soundings, which are converged using the selected clear channels, were assessed. The rejection of all IASI 182 channels also reduced the clear-sky sounding cases. In order to compare the same number of cases, the rejected soundings in clear channel selection method were substituted for the corresponding background profiles.

The results of the 1D-Var analysis for the retrieved temperature and humidity soundings are compared with the true states of the atmosphere (Figure 11). The error states of temperature and humidity in upper troposphere from the Cloudy

1D-Var analysis and the clear channel selection method are similar. We have identified that most of the selected clear channels from the clear channel selection algorithm are high-peaking channels and used to improve the upper level analysis. However, improvements of root mean square error (RMSE) in both temperature and humidity through the lower and mid-level troposphere from clear channel selection method are smaller than the Cloudy 1D-Var result. Bias in the temperature profile by the clear channel detection is also larger. This result demonstrates the limitation of the use of some low-peaking channels in the clear channel selection method. Finally, the results of the 1D-Var analysis of temperature and moisture profiles highlight the advantage that the Cloudy 1D-Var algorithm retrieves low clouds and uses more information of channels sensitive to lower troposphere.

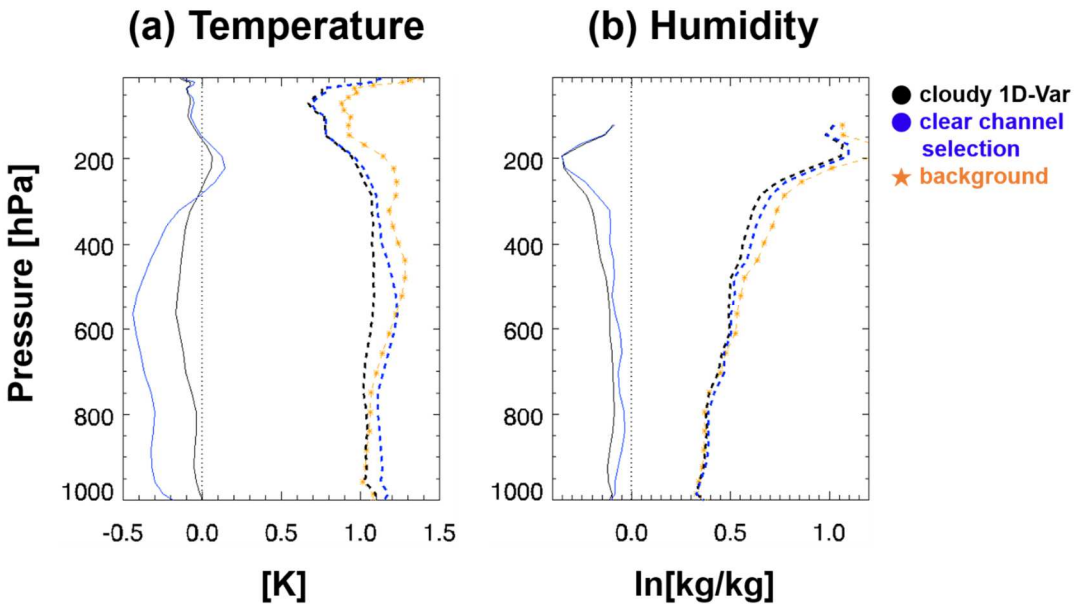


Figure 11. Mean bias (solid line) and RMSE (dashed line) of (a) temperature and (b) humidity profiles analyzed by the cloudy 1D-Var method (black) and the clear channel selection method (blue). Orange line with asterisks represents RMSE of background states.

6. Development of a new cloud parameters retrieval method

Here we introduce a new method to retrieve the first guess pair of CTP and CF (expressed as the method (b) in the dashed rectangle of Figure 2. In this study, the developed method will replace the MR method in the Cloudy 1D-Var method, in order to examine the impact of the new retrieval method on the IASI 1D-Var data assimilation. Details about the development of the retrieval method follow.

6.1. Definition of 'Optimum CTP'

This research was motivated from a theory that the first guess pair of CTP and CF, yielding a minimum residual, may not necessarily result in the best 1D-Var analysis. We define here a pair of CTP and CF yielding the best 1D-Var analysis result as 'optimum' cloud parameters. In order to obtain the optimum cloud parameters, we first take 27 pairs of CTP and CF using Eq. (6) (marked by 'x's in Figure 3) for the 27 RTTOV-pressure level in Table 2. Note that the pair of CTP and CF at the 'Δ' mark resulted in the overall minimum residual amongst 27 pairs. Each of those 27 pairs is now used as the first guess pair of CTP and CF for the IASI 1D-Var assimilation system outlined in Figure 2, and resultant temperature and humidity analysis are obtained.

In order to examine how each pair gives rise to errors in the analysis, compared to the truth, the cumulative normalized RMSE of temperature and humidity analysis is calculated by using the following equation, i.e.:

$$RMSE' = \sqrt{\frac{1}{27}\sum_{k=1}^{27}\left(\frac{T_a^k(p_c,N)-T_t^k}{\sigma_T^k}\right)^2} + \sqrt{\frac{1}{27}\sum_{k=1}^{27}\left(\frac{q_a^k(p_c,N)-q_t^k}{\sigma_q^k}\right)^2} \qquad (9)$$

where T_a and q_a are the temperature and humidity from the analysis; T_t and q_t are the truth temperature and humidity; σ_T and σ_q are standard deviations of background error for temperature and humidity, respectively. In this RMSE' calculation, the UM bottom 27 pressure levels (from surface to 102.05 hPa level) are used, and k is the index for the level.

The obtained RMSE' values for the given pairs of CTP and CF (marked by ‘x’s in Figure 3) are presented in Figure 12. It is demonstrated that the smallest RMSE' is shown at the pair (CTP = 795.09 hPa, CF = 0.44) marked by ‘∗’, instead of the pair (CTP = 436.95 hPa , CF = 0.16) marked by ‘Δ’ that showed the minimum residual. From this demonstration, it suffices to conclude that the first guess pair of CTP and CF retrieved by the MR method may not necessarily produce the smallest RMSE' in the 1D-Var analysis, suggesting that there may be a different pair of CTP and CF producing the lowest RMSE' (or best analysis

results). We call the CTP and CF pair yielding the best amongst 27 different analysis outputs as the optimum cloud parameters.

In order to construct the optimum pairs of CTP and CF from all 22,302 samples, we take the procedures summarized in Figure 13. For any given IASI radiances and corresponding background state, all possible combination of CTP and CF pairs can be introduced to the assimilation system to find a CTP and CF pair yielding the smallest RMSE'. However, because of the computational burden, we follow the procedures employed in the MR method in the UM OPS; we first find each level's CTP and CF pair throughout 27 levels using Eq. (6), and then those 27 pairs of CTP and CF (i.e. output (a) in Figure 13) are sequentially used as inputs for the following 1D-Var analysis.

For the given IASI radiances and the corresponding background atmospheric state, the 1D-Var analysis gives rise to 27 sets of profiles of temperature and humidity, which will then be compared against the truth by calculating their respective RMSE' (i.e. process (b) in Figure 13). Finally, the optical cloud parameters (CTP and CF pair) producing the smallest RMSE' result are obtained for a particular profile. By repeating the procedures summarized in Figure 13 for the entire truth data, 22,302 pairs of CTP and CF corresponding to ECMWF truth states are constructed as the optimum cloud parameters dataset, yielding the best analysis outputs. After eliminating cases either showing extreme background

atmospheric states or not yielding optimum cloud parameters, 20,327, 20,085, and 20,206 samples of truth profile, simulated IASI radiances, background states, and the optimum CTP were constructed respectively for 90°N−30°N, 30°N−30°S, and 30°S−90°S latitudinal regions.

Before optimum cloud parameters data are used for developing the new cloud retrieval method, we examine the impact of the optimum cloud estimates on the 1D-Var analysis results. Results were divided into the three latitudinal regions to examine whether results significantly vary with different latitudes. The error profiles for the 1D-Var analysis (1) using the optimum cloud parameters, and (2) using optimized CTP and CF from the MR method are given in Figure 14, along with the background error profiles. We also combined the result from the three latitudinal regions (Figure 15). It is clear that the MR method results in larger biases and higher RMSEs for both temperature and humidity for all three regions, compared with the best analysis results. This analysis result suggests that the MR method can be improved when the optimum values of CTP and CF are used.

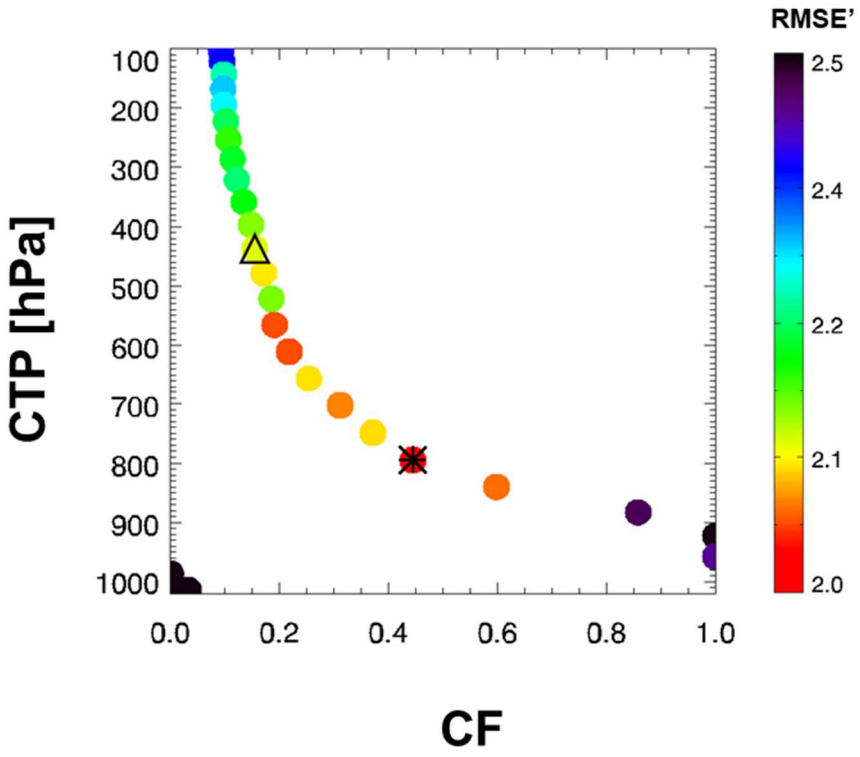


Figure 12. RMSE' at 27 pairs of CTP and CF denoted by ‘x’s in Figure 2. ‘Δ’ denotes a first guess pair of CTP and CF from the MR method whereas ‘*’ indicates the optimum CTP and CF pair showing the smallest RMSE'.

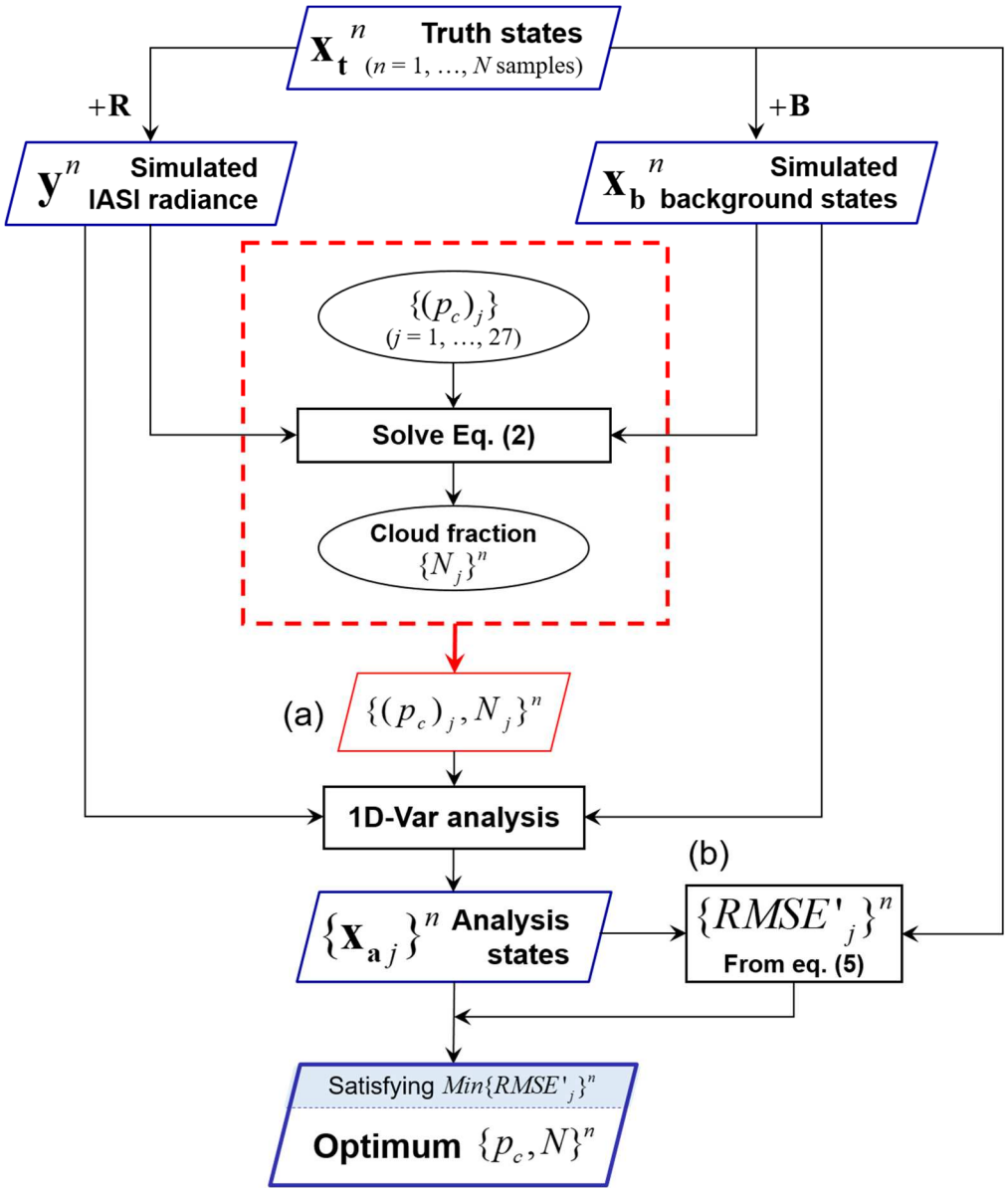


Figure 13. Schematic diagram showing the method to obtain the optimum cloud parameters. Red dashed box shows the process generating (a) 27 pairs of CTP and CF (as noted 'x' marks in Figure 2). (b) RMSE's at 27 RTTOV levels are calculated by comparing the analysis of temperature and moisture to the truth x_t, as exemplified in Figure 3, and the optimum CTP and CF values are determined. The upper index n is the ordinal number of total N samples and the lower index j represents the RTTOV pressure level. Other notations are found in the text.

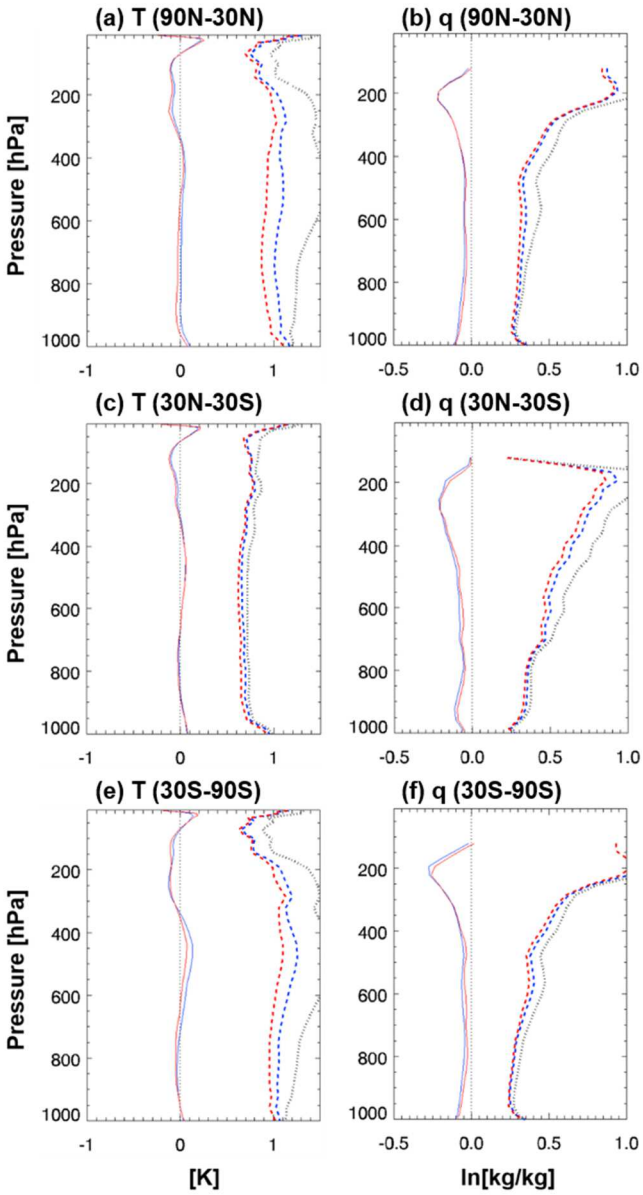


Figure 14. Mean bias (solid lines) and RMSE (dashed lines) of temperature (left panels) and humidity (right panels) profiles from the 1D-Var analysis with the use of the optimized CTP and CF from the MR method (blue) and the optimum CTP and CF (red) at (top) 90°N−30°N, (middle) 30°N−30°S, and (bottom) 30°S−90°S. Black dotted lines in the RMSE profiles represent the RMSE profiles of the background state.

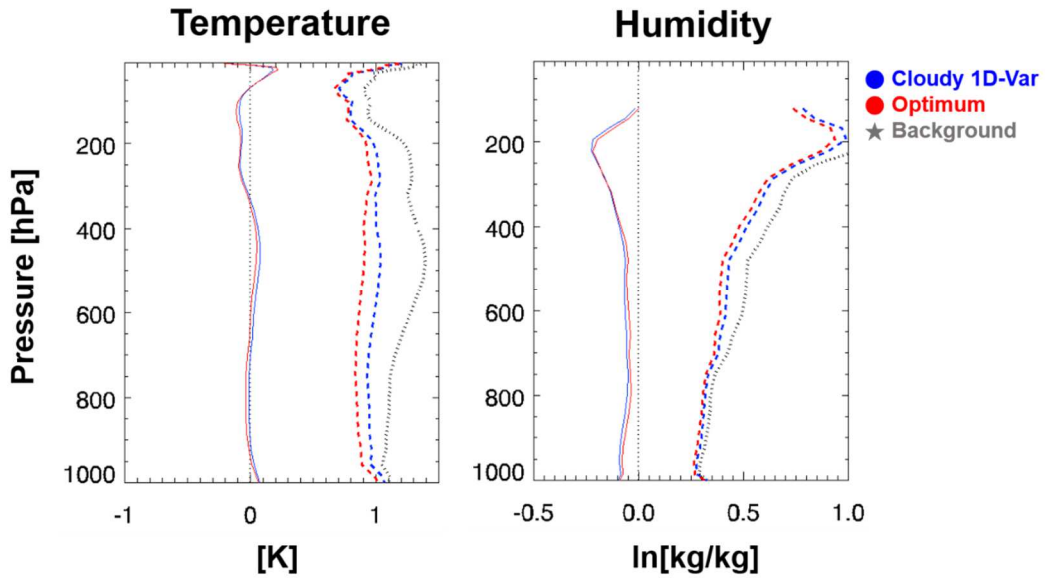


Figure 15. Same as Figure 14 except for all latitude regions.

6.2. Evaluation of original retrieval method

It is of interest to examine what factors contributed the most to the better analysis results when the optimum cloud parameters are used in the 1D-Var analysis. For this purpose, the optimum cloud parameters are compared with the optimized CTP and CF from the MR method in the three latitudinal datasets (Figure 16). In the case of CTP (Figure 16(a), (c), and (e)), there is a significant number of cases, in which the optimized CTPs from the MR method are smaller than corresponding optimum CTPs. Corresponding CF distributions are presented in terms of mean difference between the optimum CF and the optimized CF with the MR method, in order to exemplify systematic biases (Figure 16(b), (d), and (f)). It is clearly shown that CF is overestimated by the MR method if cloud top height is underestimated, in comparison to the optimum values. Conversely, underestimation of CF is clear in case of overestimated cloud top height by the MR method.

This result is consistent with the finding that the retrieved pair of CTP and CF lies along the line showing an inverse relation between CTP and CF as depicted in Figure 3 and Figure 12. Considering that overestimation of cloud top height by the MR method is much more frequent, as noted from CTP distributions (Figure 16(a), (c), and (e)), these results suggest that the MR method tends to produce

higher cloud tops with smaller cloud fractions, compared to the optimum cloud parameters.

In the IASI 1D-Var assimilation, once optimized cloud parameters are obtained, channel selection procedures are performed to find IASI channels either not contaminated by clouds or partially contaminated by clouds. Here we examine how selected channels with the optimum results differ from channels selected with MR method results. The numbers of selected channels from the optimum cloud parameters are compared with those from the MR method (Figure 17(a), (c), and (e)). Both show a large number of selected channels in the upper tropospheric CO_2 channels and less in the lower tropospheric CO_2 and window channels. The difference in the number of the selected channels is presented in Figure 17(b), (d), and (f). It is noted that the optimum cloud parameters allowed more selected channels in all three latitudinal regions, compared with the MR method. The optimum CTPs tend to be at lower altitude, giving more selected channels whose Jacobians area above the cloud top which are over 90% of the total. Moreover, because we expect not much discernible difference in the high peaking channels between two methods simply by lowering the cloud top, the increased use of high peaking channels (i.e. 1st to 30th channels) represents more converged cases by the optimum cloud parameters. These results suggest that there may be more discarded IASI channel radiances in the MR method due to the tendency to

overestimate the cloud top height, and due to the reduced convergence in the assimilation process. Compared to the MR method, those discarded channel radiances might be used in the 1D-Var analysis when the optimum cloud parameters are used.

The mean difference between observed and background radiance (O-B) is calculated for each of 182 IASI channels, and results are given for the MR method and optimum parameters in Figure 18. It is shown that O-B means in three regions are nearly the same, except that O-B means in the lower-tropospheric CO_2 and window channels (i.e. from 120 to 150 in the channel number) are slightly larger for the optimum cloud parameters. Taking this result together with Figure 17, we conclude that the optimum cloud parameters led to use of more channels in the 1D-Var analysis and gave O-B mean values similar to the MR method.

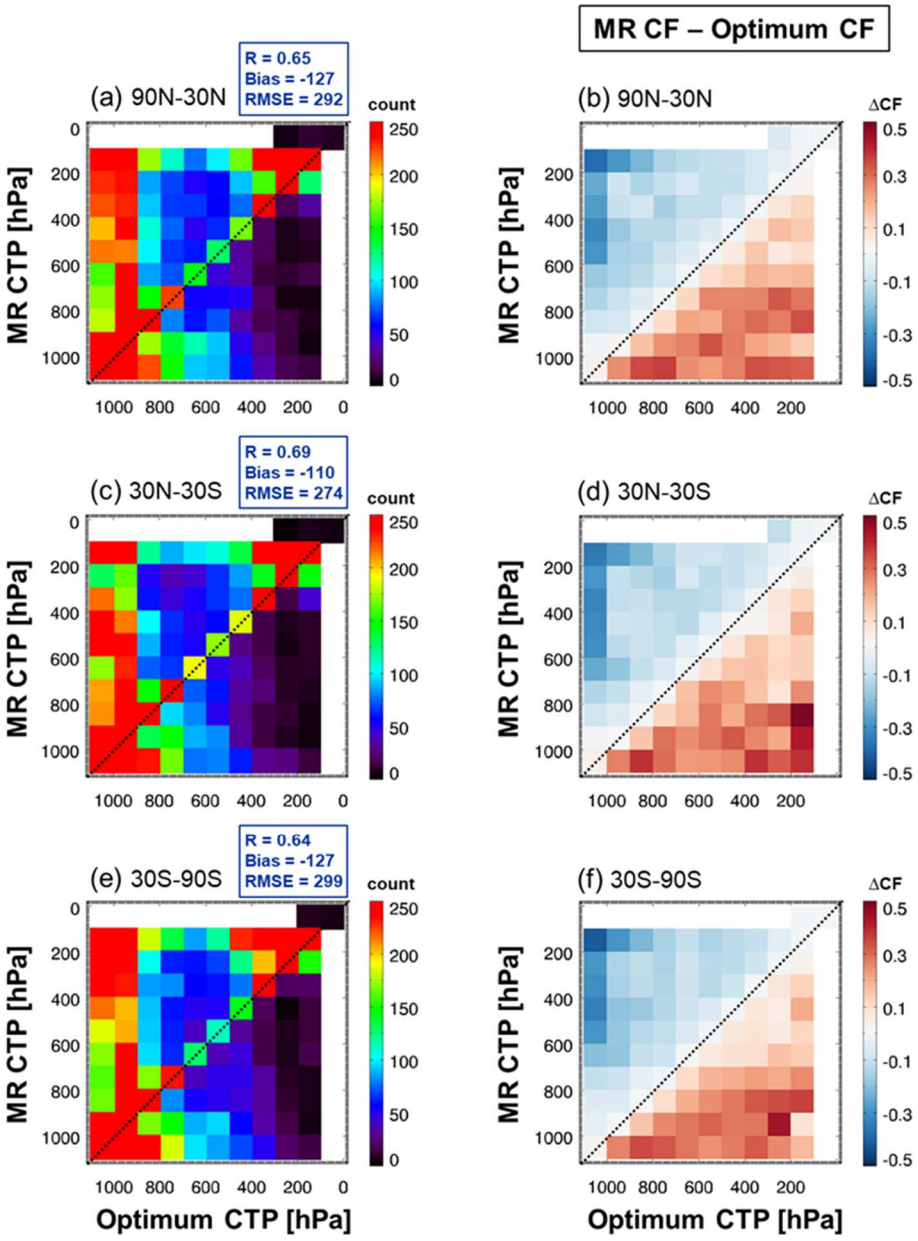


Figure 16. Two-dimensional histograms of frequencies of optimum CTP vs. optimized CTP from the MR method (left panels). Color represents data count. Mean CF differences between optimum and the MR method (right panels) are given in the same CTP pressure coordinates as in the left panels. 90°N−30°N, 30°N−30°S, and 30°S−90°S regions are given from the top to the bottom.

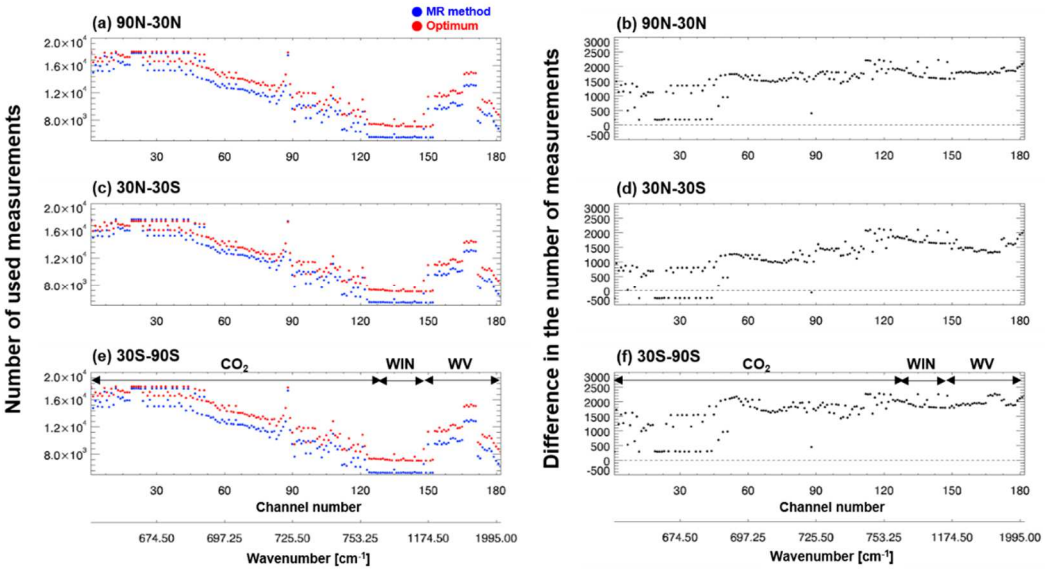


Figure 17. (left) Number of used measurements at each IASI channel from the MR method (blue) and from the optimum CTP and CF (red). (right) Difference in the number of used measurements (optimum minus MR method). Top, middle, and bottom figures represent latitudinal regions of 90°N−30°N, 30°N−30°S, and 30°N−90°S, respectively. Black arrows at bottom figures represent IASI channels in CO_2, window, and water vapor absorption bands.

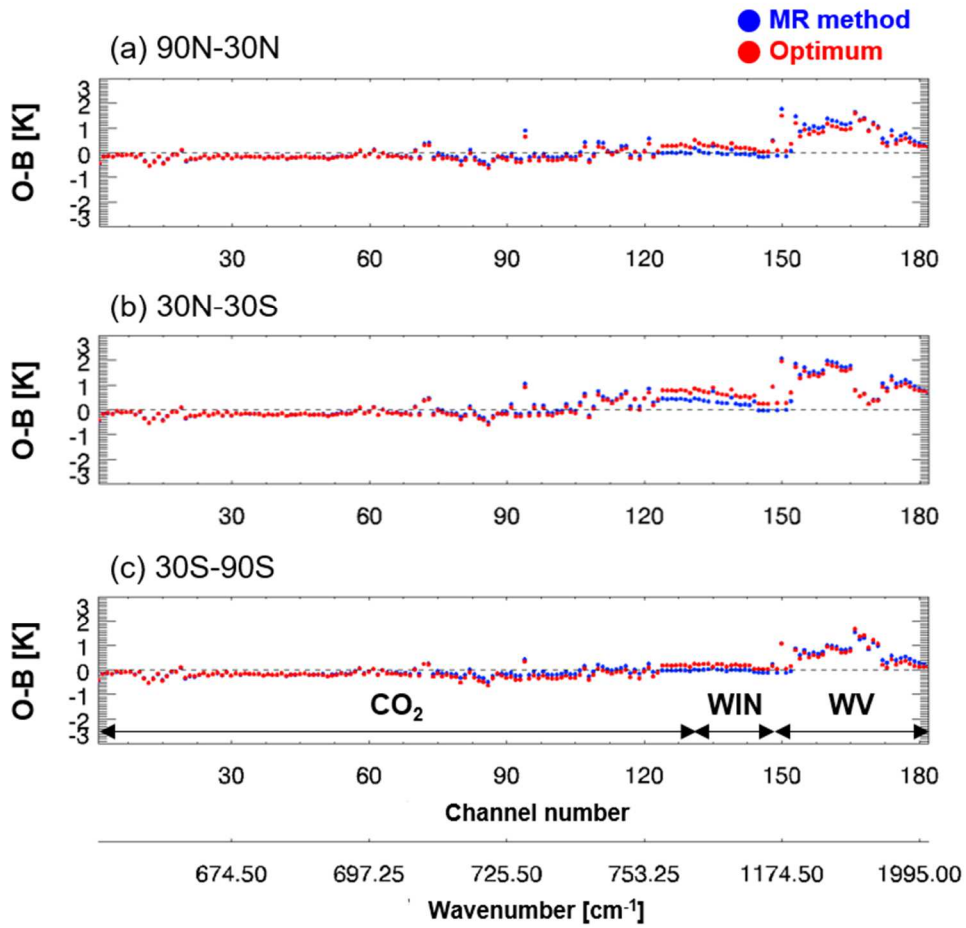


Figure 18. Mean O-B (observed minus background brightness temperature) for the MR method (blue) and the optimum CTP and CF (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands.

6.3. New retrieval method with an ANN approach

It was shown that the optimum cloud parameters produce the best 1D-Var analysis results that can be obtained from given 182 IASI channel radiances and background atmospheric states, compared to the truth states. However, the optimum cloud parameters can only be found when the truth atmospheric states are known. Thus, it is desirable to retrieve cloud parameters as close as possible to the optimum cloud parameters. In this study, we attempt to retrieve the cloud top pressure close to the optimum value by linking variables of IASI radiances and background state to the corresponding optimum cloud parameters.

In doing so, we adopt an Artificial Neural Network (ANN) approach which is well known for resolving any form of the non-linear relationship (Desai et al, 2008). We tested results with various hidden layers and neuron numbers, and empirically found that the ANN method including one hidden layer with five neurons shows the best performance of capturing the non-linearity between inputs and output used in this study.

As shown in Figure 19, the ANN model consists of three layers, i.e., input layer, hidden layer, and output layer. As inputs, 182 IASI channel radiances, background temperature and humidity profiles, and surface variables (surface temperature, skin temperature, and surface humidity) are used. The 254 neurons

in the input layer are scaled between -1 and 1, and transported to the hidden layer via synaptic weights. Five neurons in the hidden layer sum up the weighted inputs with a bias value and those 5-neuron values are passed through an activation function, which in this case is a tangent sigmoidal function. The outputs from the hidden layer are used as inputs to the output layer, and the five values produce the final output through a linear function. When the coefficients for the weights and the linear function are determined, the iterative process minimizes the RMSE of CTPs up to 145 hPa by using a backpropagation algorithm to adjust the weights appropriately (Rumelhart et al., 1986).

Amongst 20,327, 20,085, and 20,206 samples for three different latitudinal regions (90°N−30°N, 30°N−30°S, and 30°S−90°S), randomly chosen 16,293, 16,108, and 16,198 samples (about 80% of the total samples) are used for the training of the ANN model. The chosen data are further divided into 70% for training and 30% for test. The remained 4,034, 3,977, and 4,008 samples (about 20% of the total samples) are used to validate the ANN model at three latitudinal regions, respectively. It is noted that datasets allocated for training, test, and validation all show spatially and temporally homogeneous distributions (not shown). Considering that those datasets are from the randomized selection of ECMWF forecasts and show spatial and temporal homogeneity, they are considered to be independent of each other.

In the ANN model development, only CTP is determined, and then CF is obtained by inserting CTP to Eq. (6). The ANN method only determines the CTP value because CTP is more closely related to the temperature information coming from CO_2 channels, which not only compose a major portion of the 182 IASI channels but also give the largest impact on the 1D-Var analysis. Note that in the MR method, calculations of 27-level CFs (as in Eq. (6)), as well as corresponding residuals are required (as in Eq. (5)) before finding a pair of CTP and CF yielding the minimum residual. By contrast, because only one time calculation of CF is required in the ANN method (as in Eq. (6)), the computation time to find the first guess pair of CTP and CF can be reduced by the ANN method to roughly 1/54 (approximately 1.85%) of the MR method. This ANN method can replace the MR method in the flow of the IASI 1D-Var data assimilation, as depicted in Figure 2.

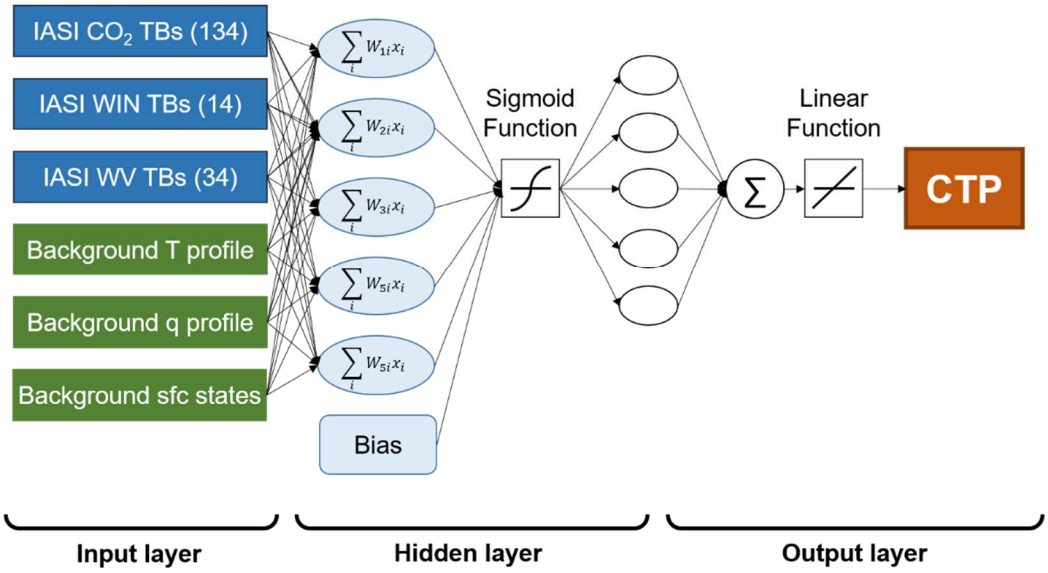


Figure 19. Structure of the ANN model showing 254 inputs (182 IASI TBs, 43-level background temperature and humidity profiles, surface temperature, skin temperature, and surface humidity), a single hidden layer with 5 neurons, and the target output (CTP).

7. Assessment of ANN retrieval method in the 1D-Var analysis

7.1. Simulation Framework

The optimized CTP and CF pairs from the MR method and the ANN method are compared with the optimum CTP and CF validation dataset of 12,019 samples (Figure 20). The optimized CTPs from the MR method show a bias of -119 hPa, RMSE of 290 hPa, and a correlation coefficient of 0.66 with the optimum CTPs. Negative bias suggests that the optimized CTPs from the MR method are likely smaller (i.e. higher cloud top) than the optimum CTPs. The density histogram of CTP for the ANN method (Figure 20(c)), on the other hand, shows better agreement with the optimum CTPs, with 0.87, -19 hPa, and 169 hPa for correlation coefficient, mean bias, and RMSE, respectively.

For the CF comparison, results are presented in the form of the mean difference from the optimum CF, in the same CTP coordinates as in Figure 20(a) and (c). The results showing generally higher cloud tops tend to have smaller cloud amounts. Opposite behaviors are found in the case of underestimates of the cloud top height. The general patterns shown in the difference between the ANN method and the optimum approach (Figure 20(d)) are similar to results from the MR method (Figure 20(b)). However, considering that the majority of frequencies of

CTPs from the ANN method are located along a diagonal line (Figure 20(c)) where the CF difference appears to be small, CFs from the ANN method should be much more similar to the optimum CFs. In conclusion, biases in both cloud top height and cloud fraction shown in the MR method are substantially removed by the ANN method, resulting in CTPs as well as CFs which are closer to the optimum values.

It is noted that the use of optimum cloud parameters yields the best 1D-Var results under the assumption of the single-layer grey-body cloud. Therefore, we expect that assignment of CTP and CF closer to the optimum parameters using the ANN method would give better analysis results (i.e. temperature and humidity profiles). To test this hypothesis, a 1D-Var analysis (shown in the second stage of the flow in Figure 2) was performed for all 12,019 validation cases. The resultant statistics are from all cases regardless of the convergence state, including non-converged cases which are replaced by corresponding background profiles. Convergence statistics are given in Table 3, in which all 4,214 cases for clear 1D-Var and 7,805 cases for cloudy 1D-Var analysis are found among 12,019 cases for the optimum cloud parameters. Non-convergence cases are not shown in the optimum cases, because the optimum cloud parameters were chosen only if the convergence criteria were met in the 1D-Var analysis. From the MR method, 3,211 and 7,340 cases were found to be converged in clear and cloudy 1D-Var, respectively. The remaining 1,468 cases were not converged and thus were

replaced by their background state. Meanwhile, from the ANN method, 3,516 and 7,547 cases for clear and cloudy 1D-Var were converged, with 956 non-converged cases replaced by their background state. These statistics suggest that the ANN method yields more converged cases in the 1D-Var analysis. As the ANN method gives CTPs closer to their optimum values, retrieved CFs are expected to become closer to their optimum values as well, according to Eq. (6). Conversely, better cloud retrievals mean better clear scene determination, which may lead to more converged cases over clear skies.

Channel selection for the 1D-Var analysis was performed by using the optimized cloud parameters from the MR method and the ANN method, and by using the optimum cloud parameters for the 12,019 validation samples (Figure 21). It is shown that more channels are used by the ANN approach over CO_2, water vapor, and window bands, compared to the channels selected by the MR method (Figure 21(a)). The ANN method appears to allow the 1D-Var analysis to use more cloud-affected channels because of the larger number of converged cases in both clear and cloudy 1D-Var, compared to the MR method. In particular, lower cloud tops by the ANN method induce higher cloud amounts, which likely enable cloud-affected channels to be more available in the 1D-Var analysis. Considering that the cloudy-sky data assimilation is about to use channels not contaminated by cloud

presence, more upper-level temperature sensitive channels can be available in the 1D-Var analysis because of the lower cloud tops.

The O-B means of selected channels by three methods are given in Figure 21(b). It shows that all three methods gave nearly the same O-B means for CO_2 channels except lower-level picking CO_2 channels (i.e. from 120 to 135 in the channel number). The O-B mean values in the window channels show weak positive biases for the ANN method and the optimum cloud parameters. Overall, it is noted that the ANN method accommodates more cloud-affected channels than the MR method while results in quite similar O-B means.

The error statistics of the 1D-Var analysis from the MR method, the ANN method, and the optimum cloud parameters are shown in Figure 22. It is clear that the ANN method resulted in reduced biases in both temperature and humidity profiles, compared to results from the MR method. The temperature RMSE was also improved by the ANN method throughout the entire layer, particularly in the mid-troposphere between 400 and 700 hPa. The RMSE profiles by the ANN method tends to be closer to those from the use of the optimum cloud parameters. However, the error profiles of humidity show that the ANN method gave RMSE profiles hardly discernible from the results from the MR method, although reduced bias by the ANN method is noted. The smaller improvement in the water vapor field by the ANN method may be due to the smaller portion of the used water

vapor channels in the 1D-Var analysis. Relatively larger O-B over the water vapor band might also contribute to such smaller improvement. This seems to be characteristics of the current IASI 1D-Var assimilation, in which the temperature signal from the CO_2 channels is heavily weighted, due to the larger uncertainty of the background water vapor fields (Hilton et al, 2012). By the same reasons, significant differences are not shown in the humidity field even if the optimum cloud parameters are used. Nevertheless, it can be concluded that the ANN method tends to improve the error statistics in the 1D-Var analysis by selecting more channels, compared to the MR method.

Since ANN coefficients are dependent upon latitudinal regions, 4,034, 3,977, and 4,008 validation samples are used to validate the ANN model at 90°N−30°N, 30°N−30°S, and 30°S−90°S regions, respectively. But, because region-dependent results are similar to each other, we provide individual results in the Figure 23, Figure 24, and Figure 25.

Table 3. Number (percentage) of converged and non-converged cases from three methods in the 1D-Var analysis using the simulation dataset.

Method	Converged cases (cloudy 1D-Var)	Converged cases (clear 1D-Var)	Non-converged cases
Optimum CTP&CF	7,805 (65%)	4,214 (35%)	-
MR method	7,340 (61%)	3,211 (27%)	1,468(12%)
ANN method	7,547 (63%)	3,516 (29%)	956 (8%)

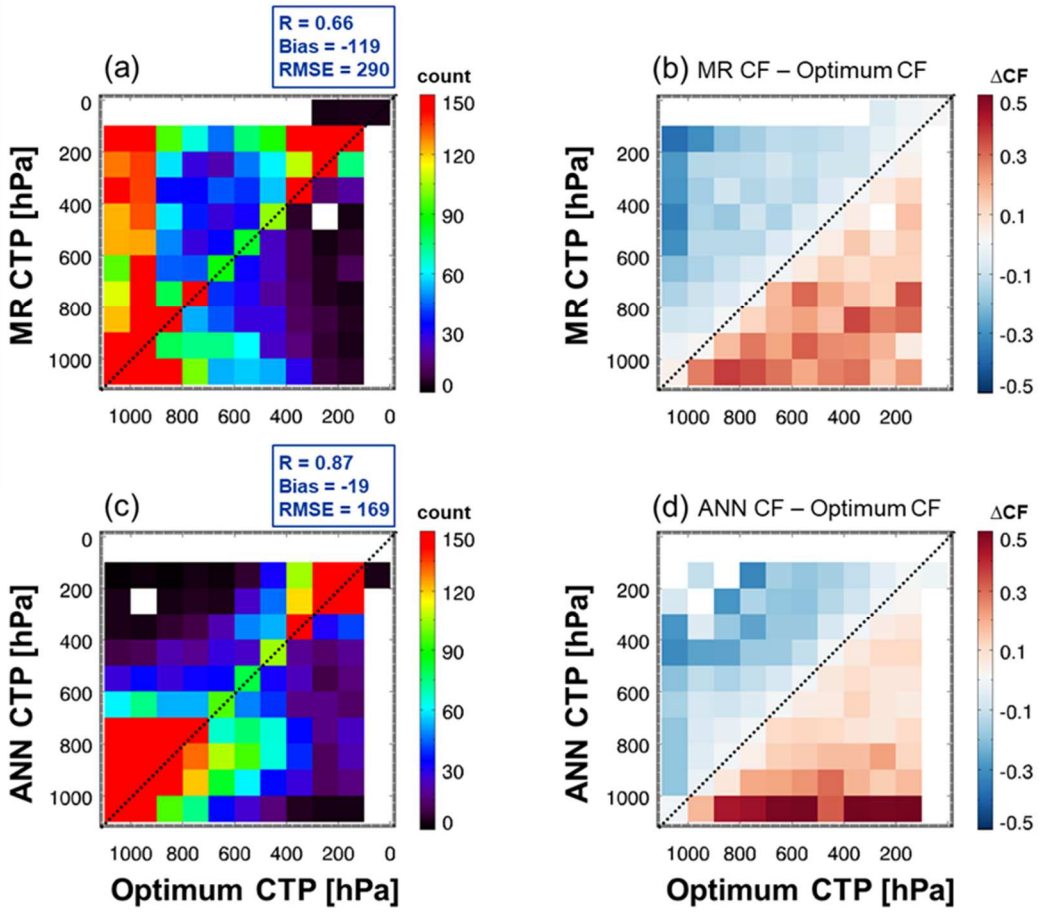


Figure 20. (left) Two-dimensional histograms of frequencies of (a) optimum CTP vs. optimized CTP from the MR method (MR CTP), and (c) optimum CTP vs. optimized CTP from the ANN method (ANN CTP). Color represents data count. (right) Mean CF difference (b) between optimum and the MR method, and (d) between optimum and the ANN method. The mean differences are given in the same CTP pressure coordinates as in the left panels.

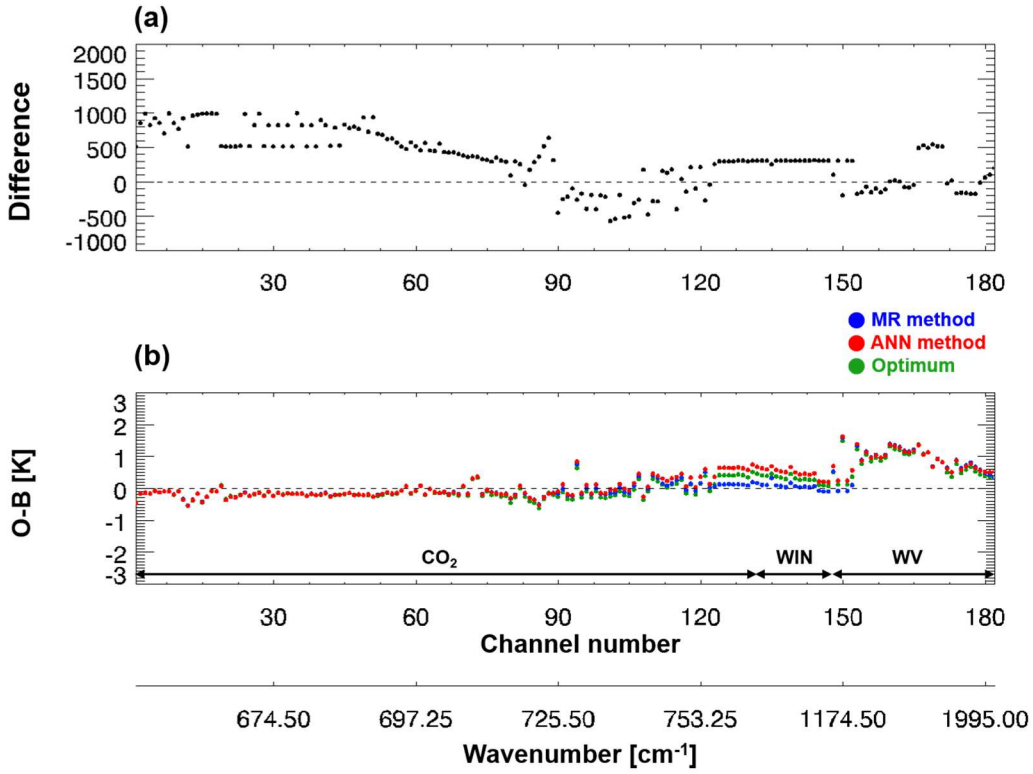


Figure 21. (a) Difference in the number of used measurements (ANN method minus MR method) at each of the 182 IASI channels. (b) Mean O-B for the optimum (green), the MR method (blue), and ANN method (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands.

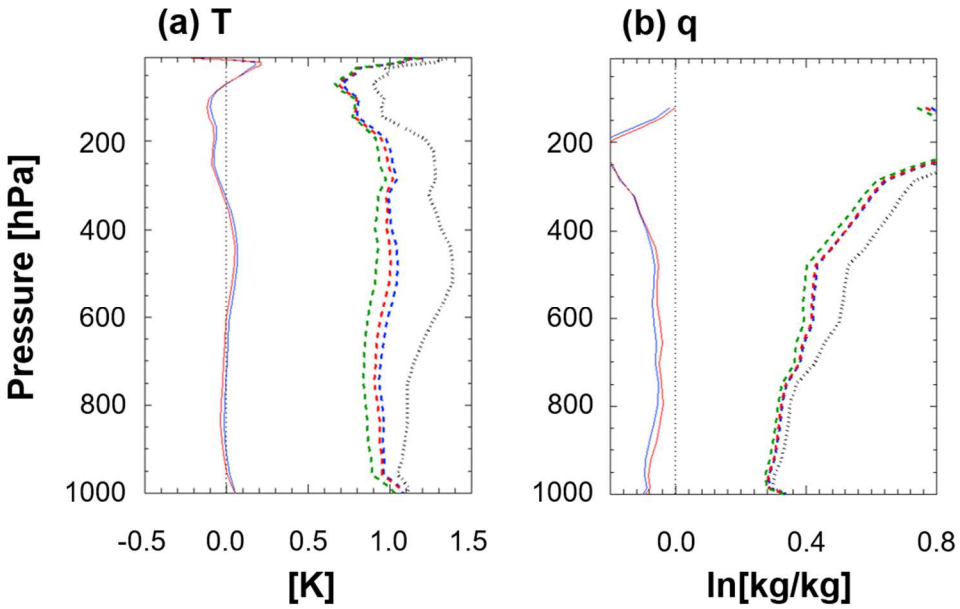
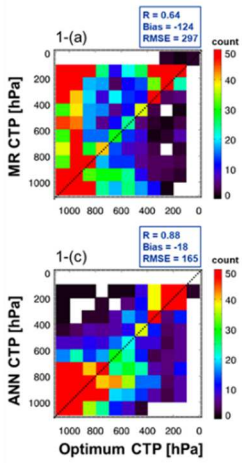
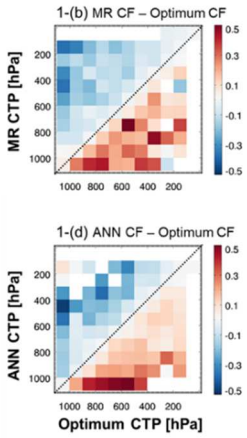


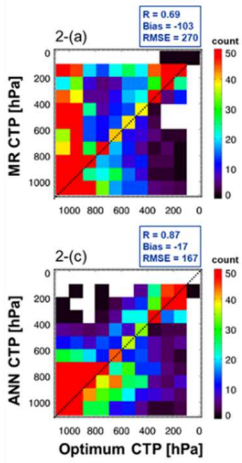
Figure 22. Mean bias (solid lines) and RMSE (dashed lines) of (a) temperature and (b) humidity analysis profiles from the 1D-Var analysis with the use of the MR method (blue), and ANN method (red), and the optimum values (green). Black dotted lines in the RMSE profiles represent the RMSE profiles of the background state.

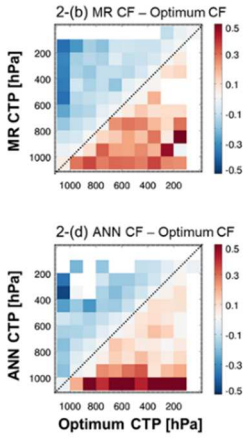
90°N-30°N



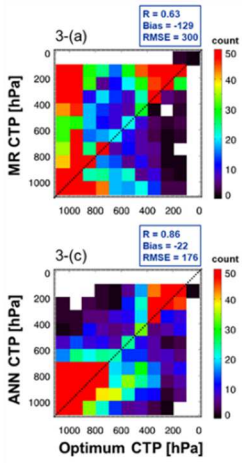


30°N-30°S





30°S-90°S



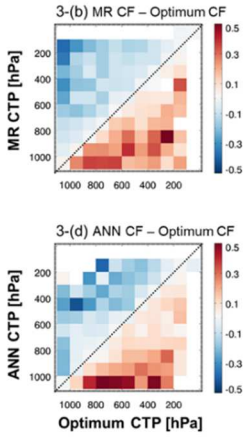
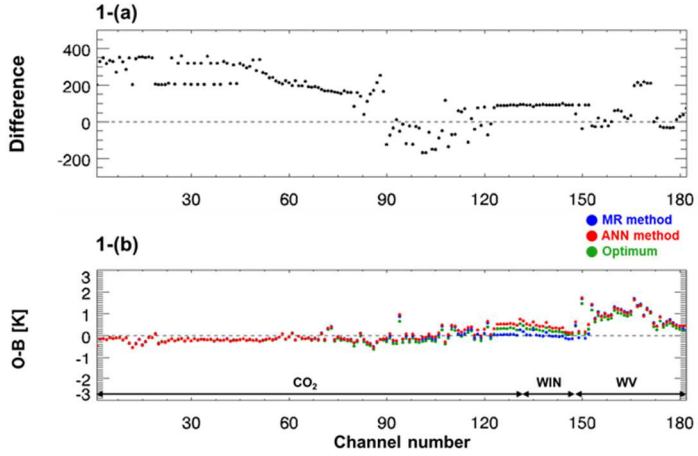
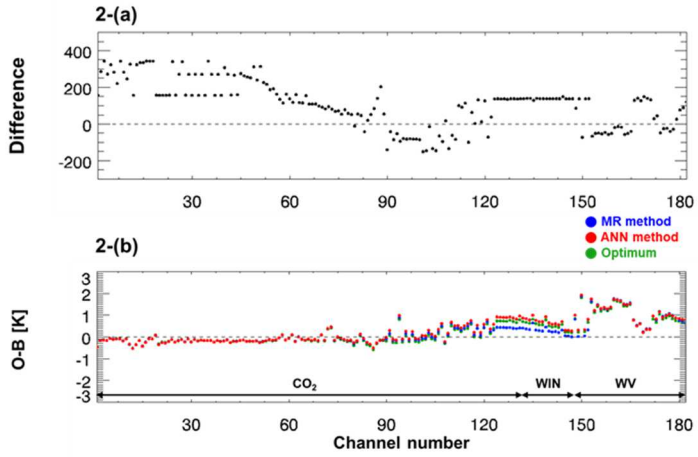


Figure 23. (Left) Two-dimensional histograms of frequencies of (a) optimum CTP vs. optimized CTP from the MR method (MR CTP), and (c) optimum CTP vs. optimized CTP from the ANN method (ANN CTP). Color represents data count. (right) Mean CF difference (b) between optimum and the MR method, and (d) between optimum and the ANN method. The mean differences are given in the same CTP pressure coordinates as in the left panels. The results for the ANN model in (1) 90°N−30°N, (2) 30°N−30°S, and (3) 30°S−90°S regions are given from the top to the bottom.

90°N-30°N



30°N-30°S



30°S-90°S

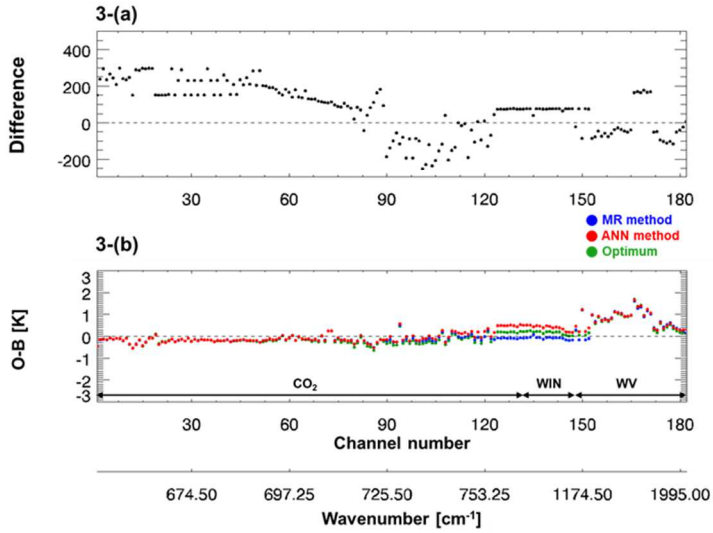
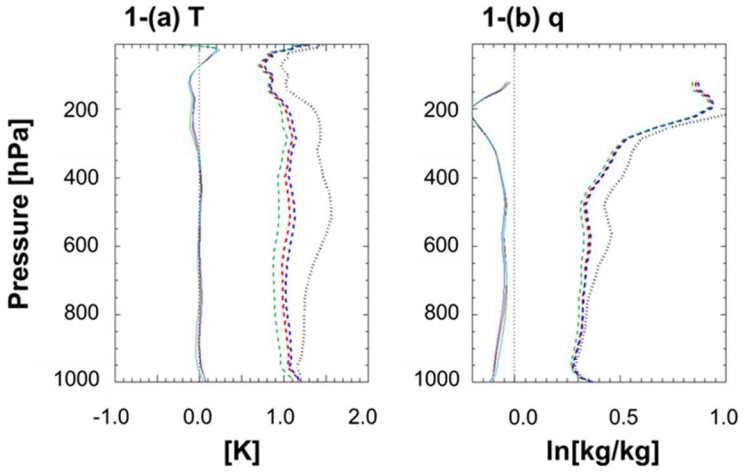
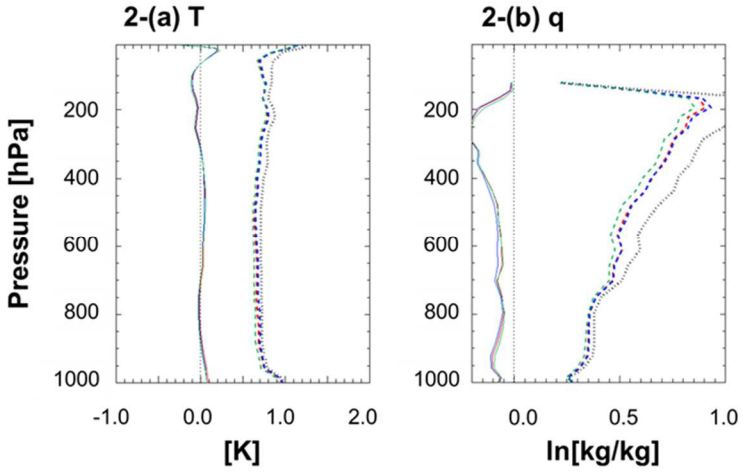


Figure 24. (a) Difference in the number of used measurements (ANN method minus MR method) at each of the 182 IASI channels. (b) Mean O-B for the optimum (green), the MR method (blue), and ANN method (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands. The results for the ANN model in (1) 90°N−30°N, (2) 30°N−30°S, and (3) 30°S−90°S regions are given from the top to the bottom.

90°N-30°N



30°N-30°S



30°S-90°S

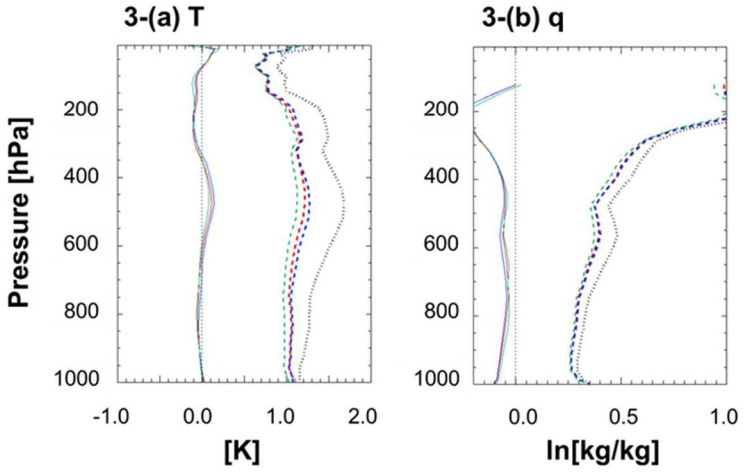


Figure 25. Mean bias (solid lines) and RMSE (dashed lines) of (a) temperature and (b) humidity analysis profiles from the 1D-Var analysis with the use of the MR method (blue), and ANN method (red), and the optimum values (green). Black dotted lines in the RMSE profiles represent the RMSE profiles of the background state. The results for the ANN model in (1) 90°N−30°N, (2) 30°N−30°S, and (3) 30°S−90°S regions are given from the top to the bottom.

7.2. Experiments with the UM NWP system

In order to examine the impact of the ANN method in the UM weather forecasts, we implemented the ANN method in the UM OPS. Two global model assimilation trials including control and experiment runs were conducted for a period from 15 July to 13 August 2017. The UM model version 10.2 at a resolution of 25 km was used, and the Variational Bias Correction (VarBC) scheme described in Auligné et al. (2007) was applied in both trial experiments. Here, the trial run with the MR method is referred to as the control run while the trial run with the ANN method is referred to as the experiment run.

The optimum CTPs were obtained in the simulation framework because RMSEs of temperature and humidity profiles can be calculated against the given truth profiles. However, it is practically impossible to define optimum cloud parameters in the operational data assimilation system, because the true atmospheric states are not known. Thus, we simply compare the optimized CTPs from the MR method and the ANN method, in order to confirm whether the CTP results are consistent with the results from the simulation framework. In Figure 26(a), the scattergram of CTPs from the UM OPS is shown for 00 UTC 30 July 2017, which is from 109,155 samples over open ocean. It is shown that a majority of CTPs retrieved by the ANN method are lower than CTPs from the MR method. Although these UM OPS results appear to be consistent with results from the

simulation dataset, the MR method produces cloud tops excessively higher than those from the ANN method. Because of the higher cloud tops, the MR method should have produced smaller CFs, as demonstrated in Figure 27(b).

In Table 4, among the total 109,155 samples at 00 UTC on 30 July 2017, converged cases for clear and cloudy cases are 24,626 and 54,794, respectively, in the control run, whereas 13,944 and 66,251 cases in the experiment run were converged. Therefore, a total of 79,420 and 80,195 scenes were used in the control run and the experiment run, respectively. It is interesting to note that converged clear-sky cases are more in the control run (i.e. MR method), in contrast to the result that both converged clear-sky and cloudy-sky cases are more for the ANN method in the simulation framework. This may be due to the bias correction process in the UM OPS, which will affect the IASI radiances and background states used in the UM OPS. As noted in Figure 26(b), the MR method tends to yield substantially less cloud amounts when pre-processed IASI radiances and background states are used, likely leading to more clear-sky scenes employed in the assimilation. Overall, we have identified that more cases converged by the ANN method (corresponding to an increase of about 1% compared with the MR method), due to more cloudy scenes used in the data assimilation with the ANN method.

We also examine the difference in the number of selected IASI channels and in the O-B means of used channels in the data assimilation between two trials, using the same 109,155 samples (Figure 27). The use of the ANN method in the UM OPS resulted in more CO_2 channels for the data assimilation process is shown in Figure 27(a), compared to the MR method results. However, the O-B means over the CO_2 band appear similar to each other between the two methods (Figure 27(b)). This is consistent with the result from the simulation framework (Figure 21) which showed a common feature that the ANN method used more cloud-affected channels by lowering the cloud top in spite of the same O-B means. On the other hand, the ANN method resulted in less window channels used and larger negative O-B means, compared to the MR method results. It may be due to the tendency that the ANN method produces lower cloud tops and larger cloud fractions. Under the assumption of the grey-body single layer cloud, the same radiance can be interpreted as lower cloud top and larger cloud fraction, instead of higher cloud top and smaller cloud fraction. Therefore, the ANN method probably assigns the clear-sky scenes that were determined by the MR method as cloudy scenes, which effectively prevents the use of window channels.

In order to assess the impact of the ANN method on the initial analysis field of the UM operational assimilation system, we examined analysis fields collected during the two-week period from 30 July to 13 August 2017. More than 535 global

daily radiosonde observations were used over the 14 day period and errors were calculated by comparing analysis fields with the radiosonde observations. In the operational UM OPS, instead of performing the 1D-Var analysis process, the selected IASI channels are passed to the 4D-Var assimilation. Thus, in this study, the analysis results at the T+0 forecast time from the 4D-Var system were evaluated, to examine the 1D-Var analysis results from each scheme in the UM OPS. Figure 28 shows RMSEs of temperature and relative humidity of analysis at the T+0 UM global forecast time. The experiment run shows RMSEs in temperature and relative humidity analysis profiles were slightly reduced, compared to the control run. The RMSE results of temperature and relative humidity for each latitudinal band are shown in Figure 29 and Figure 30, respectively. It can be seen that the temperature RMSEs in the southern hemisphere, and the relative humidity RMSEs in all latitude bands are reduced in the experiment run. However, improvement in the initial fields in the UM forecasts by the ANN method appears to be rather minor, leading to a conclusion that the overall impact on global weather forecasting of the ANN method employed in the assimilation system is found to be neutral.

Table 4. Number (percentage) of converged and non-converged cases from the MR and ANN methods in the 1D-Var analysis of the UM OPS (00 UTC 30 July 2017).

Method	Converged cases (cloudy 1D-Var)	Converged cases (clear 1D-Var)	Non-converged cases
MR method	54,794 (68%)	24,626 (31%)	1,173(1%)
ANN method	66,351 (82%)	13,944 (17%)	398 (1%)

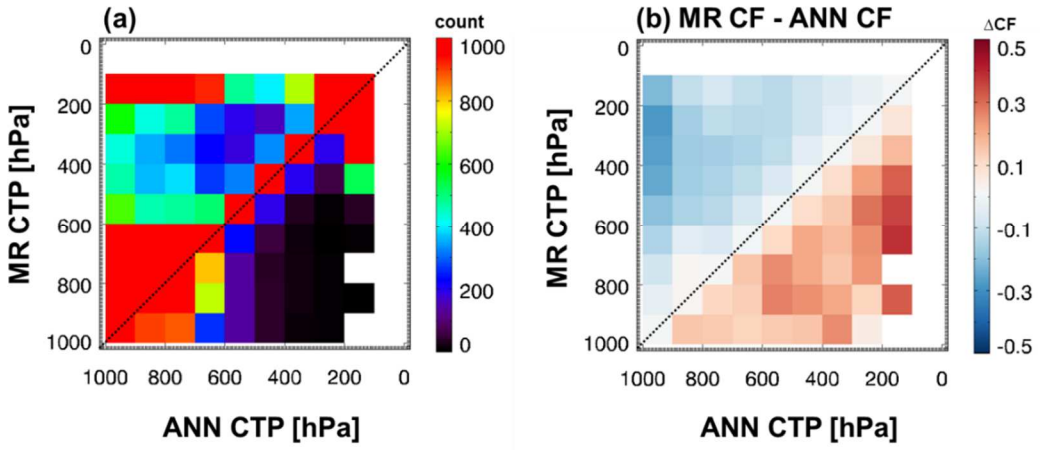


Figure 26. (a) Two-dimensional histograms of frequencies of ANN CTP vs. MR CTP in the UM OPS system. Color represents data count. (b) Mean difference between MR CF and ANN CF in the UM OPS system. The mean differences are given in the same CTP pressure coordinates as in the left panels.

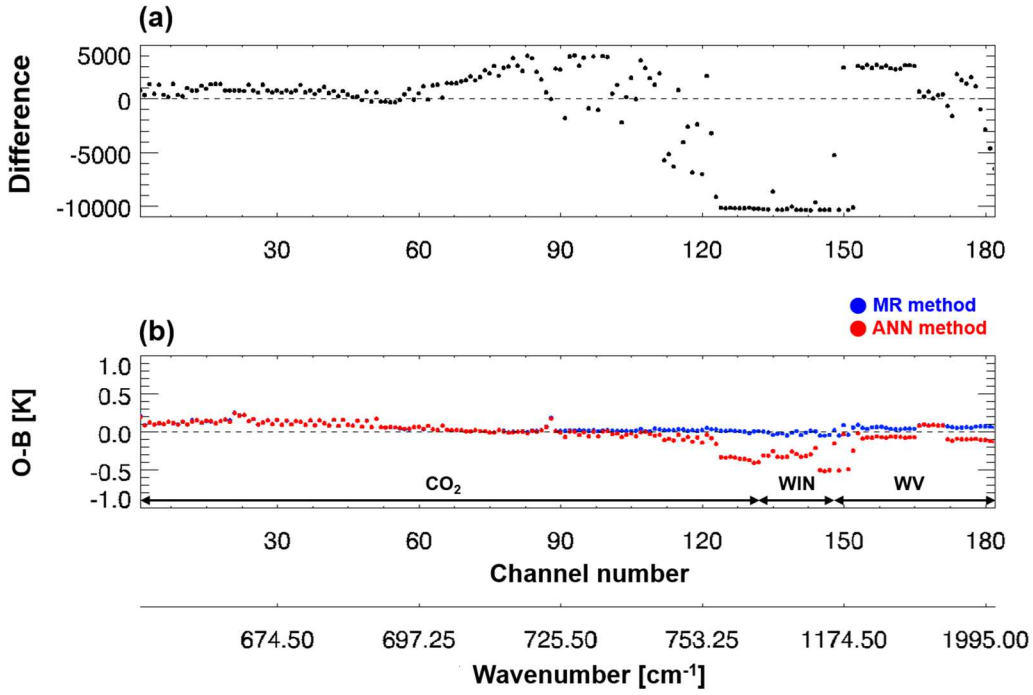


Figure 27. (a) Difference in the number of used measurements (ANN method minus MR method) at each of the 182 IASI channels. (b) Mean O-B for the optimum (green), MR method (blue), and for ANN method (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands.

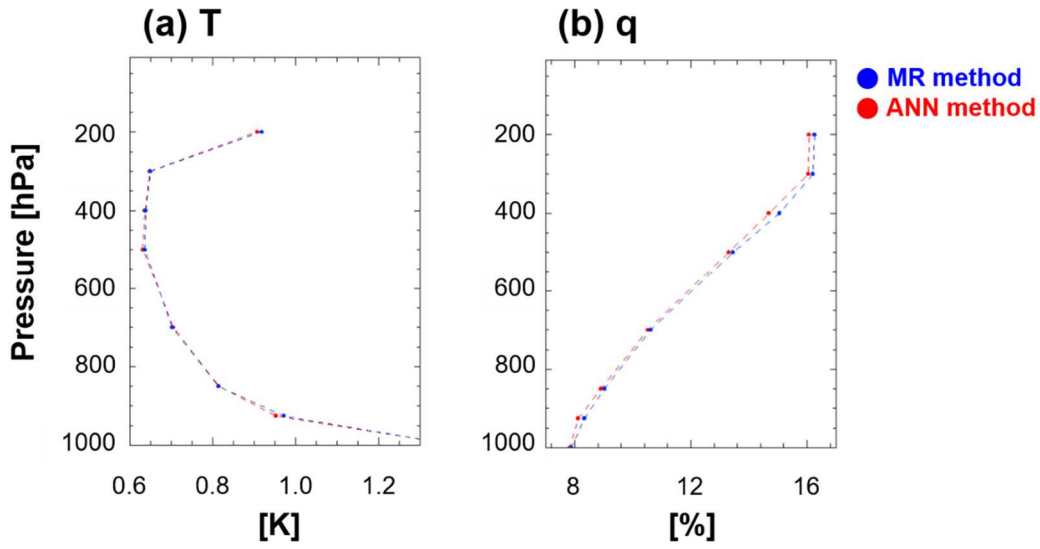


Figure 28. RMSE profiles of (a) temperature and (b) relative humidity in the UM OPS analysis with the MR method (blue) and ANN method (red).

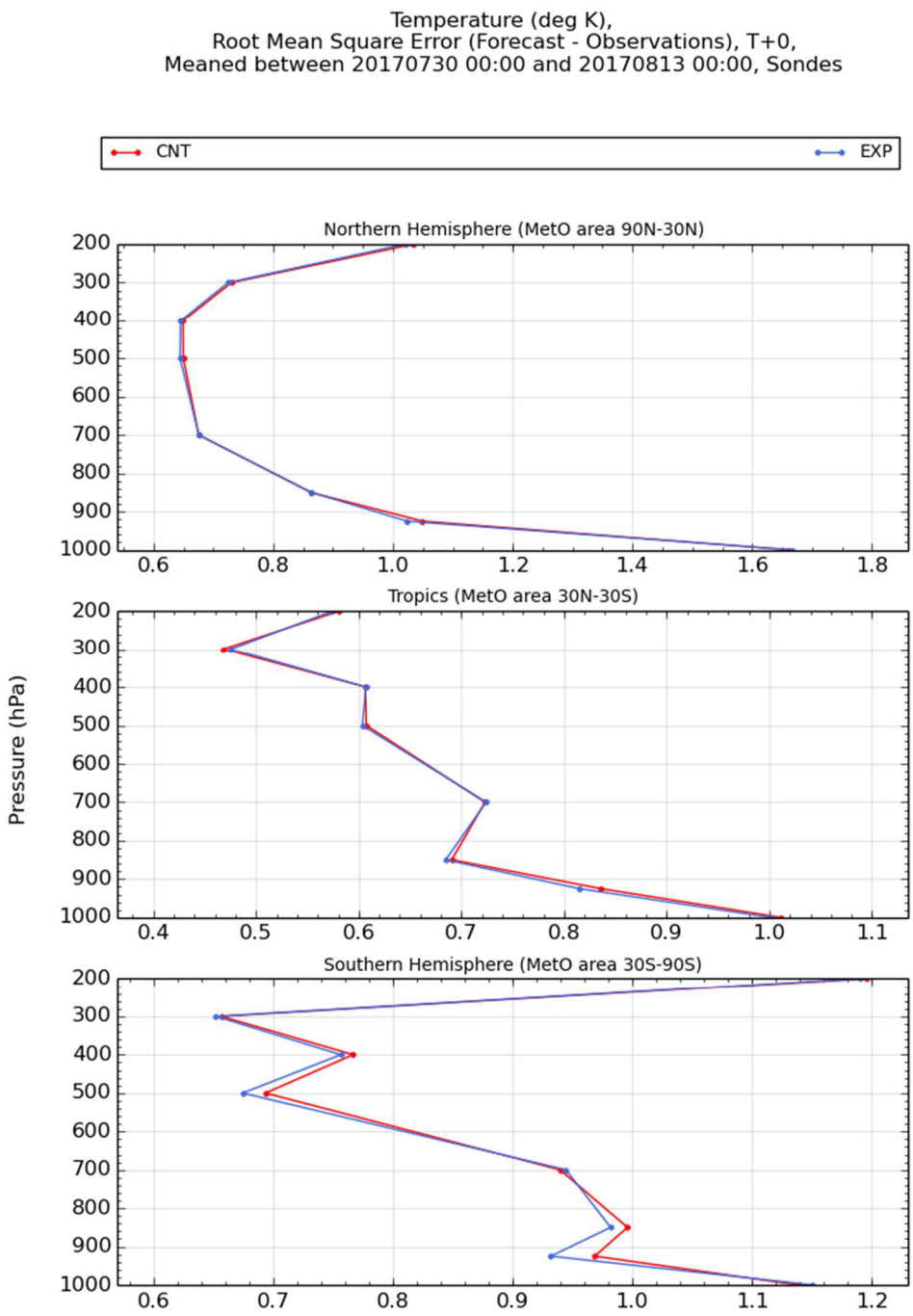


Figure 29. RMSE for temperature profiles at T+0 in the control (red) and the experiment (blue) runs at (a) 90°N−30°N, (b) 30°N−30°S, and (c) 30°S−90°S.

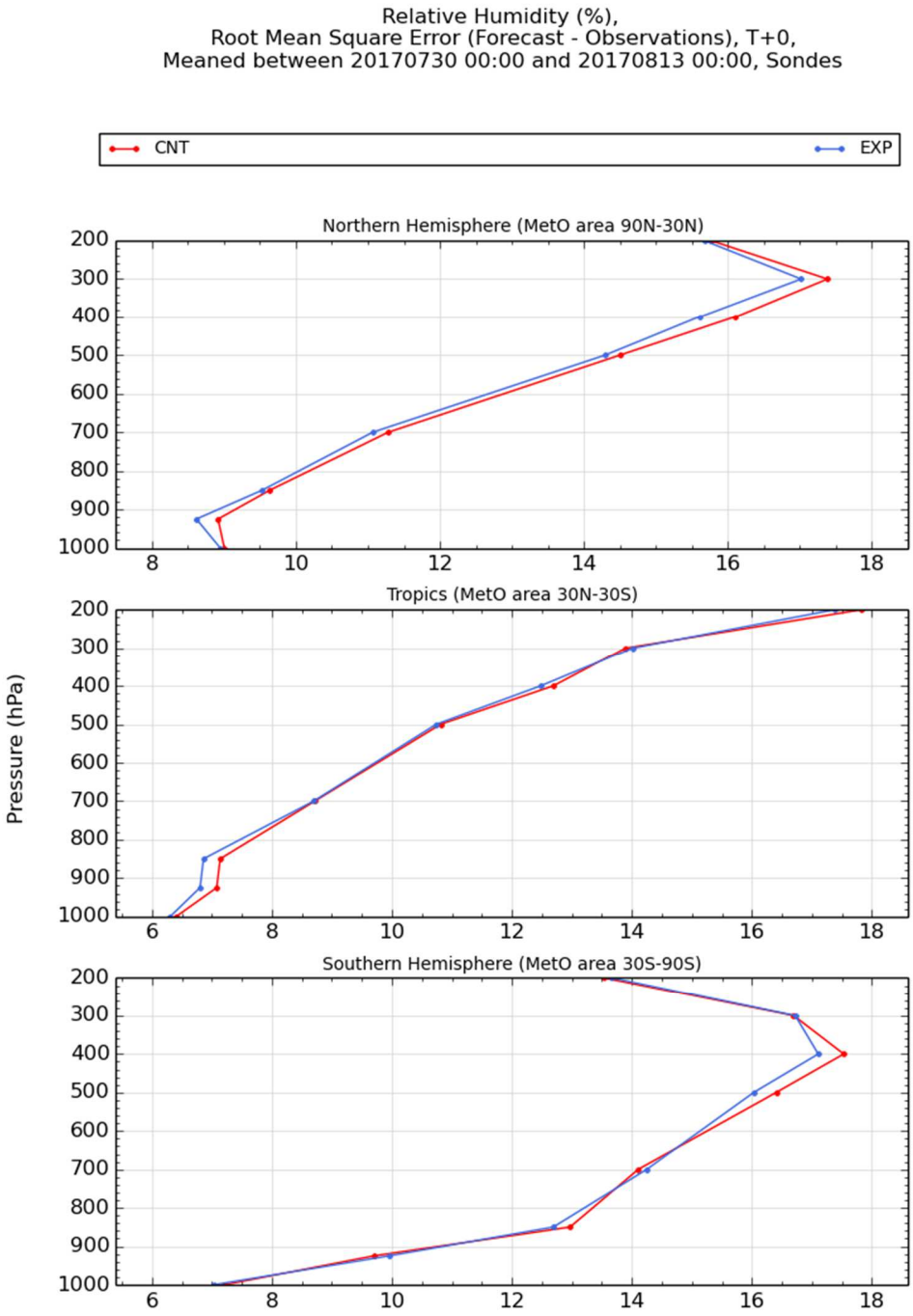


Figure 30. RMSE for relative humidity profiles at T+0 in the control (red) and the experiment (blue) runs at (a) 90°N−30°N, (b) 30°N−30°S, and (c) 30°S−90°S.

8. Impact study of ANN method on the UM forecast

In the simulation framework, the optimum cloud parameters making the best 1D-Var analysis results were obtained in correspondence with the RTTOV-simulated IASI radiances and the simulated background states. The ANN method was trained to make these optimum cloud parameters using the RTTOV-simulated IASI radiances and the simulated background states as inputs. In fact, the previous results validated that the newly developed ANN method obtains CTP closer to corresponding optimum values leading to better temperature and moisture analysis. However, when the ANN method was implemented in the UM OPS, more channel information was used only over the CO_2 and water vapor channels and the number of used channels was reduced in the channels sensitive to the lower tropospheric atmosphere. Thus, it was hard to find significant improvement from the ANN method in the UM OPS, compared to the temperature and moisture analysis from the MR method.

The biggest difference between the results in the simulation framework and those in the UM OPS is the convergence rate in the clear 1D-Var. The converged cases in the clear 1D-Var were increased when the MR method was used in the UM OPS, compared to the ANN method. In most of these increased cases, the CTPs were retrieved higher than 200 hPa by the MR method. Only a small amount

of channels, which are sensitive to upper-tropospheric atmospheric states, is selected when the retrieved cloud top is in the upper troposphere and the scene is considered to be cloudy. On the other hand, when the retrieved cloud top is the same but the scene is considered to be clear, all 182 channels are used. Considering that the simulation of the cloudy-sky infrared radiances for the upper-level clouds has large errors, these increased clear 1D-Var cases are likely to use cloud-contaminated channels. Thus, we excluded the cases having retrieved CTP smaller than 200 hPa to see the further impact of the ANN method on the UM global forecast.

Two global model assimilation trials including control and experiment runs were conducted for the same period from 15 July to 13 August 2017 as section 7.2. The UM model version 10.2 at a resolution of 25 km was used, and the VarBC scheme described in Auligné et al. (2007) was applied in both trial experiments. Here, the trial run with the MR method is referred to as the control run while the trial run with the ANN method is referred to as the experiment run, again. In both trial runs, the scenes having the retrieved CTP smaller than 200 hPa were rejected and not used in the data assimilation process.

8.1. Assessment of experiments in the UM NWP system

Before looking at the impact of the ANN method on the UM global forecast, we tried to see whether the same results from the experiment in section 7.2 were obtained in the assimilation experiment excluding the scenes having the cloud higher than 200 hPa level. The retrieved CTPs in the control and experiment runs were compared in Figure 31. It can be also seen that the ANN method produces lower cloud than the MR method. Table 5 also shows the number of converged cases in each method among the total 107,681 samples at 00 UTC on 30 July 2017. 29,782 and 69,511 cases were converged for clear and cloudy cases, respectively, in the control run, whereas 19,874 and 80,937 cases in the experiment run were converged. Therefore, a total of 99,293 and 100,811 scenes were used in the control run and the experiment run, respectively. For both clear and cloudy cases, converged cases were increased in the experiment run. Accordingly, the ANN method makes more channel information available in all channel bands as shown in Figure 32. Figure 32(b) confirms that more selected IASI channels in the experiment run still have similar O-B values as in the control run.

The initial analysis fields of the UM 4D-Var had been verified using radiosonde observation to assess the performance of the ANN method in the UM OPS. However, radiosonde observations are limited over the ocean area, where the

cloud retrieval algorithms are actually used. In order to make the same density of data used for the analysis verification over the ocean and land, it will be essential to assess the 4D-Var analysis field at T+0 with respect to reanalysis field. Using the same density of the verification data in both ocean and land regions, it would be expected to see more detailed impact from the ANN method. Thus, in this study, the ECMWF reanalysis 5th generation (ERA5) hourly data on 37-pressure level was used to verify the UM short-term forecast field of temperature and moisture from T+0 to T+72 produced by each control and experiment run.

Figure 33 presents the bias and RMSE of the UM global forecast of temperature profiles from T+0 to T+72 forecast time in each experiment, by using ERA5 reanalysis data as a reference. Both bias and RMSE profiles show that the forecast errors in the experiment run is similar to that in the control run at the whole forecast time. The UM global forecasts of humidity profiles from T+0 to T+72 forecast time were also verified in Figure 34. The RMSE of lower-level moisture at T+0 forecast time in the experiment run is slightly higher, but the errors in the experiment run are smaller in the upper layer. As the forecast time increases, the error statistics become the same in both experiments. It can be found that the impact of the ANN method on the UM global temperature and moisture forecasts is neutral even when the ANN method is implemented on the lower cloud cases in the UM OPS.

Table 5. Number (percentage) of converged and non-converged cases from the MR and ANN methods in the 1D-Var analysis of the UM OPS, only for the cases having retrieved CTP smaller than 200hPa (00 UTC 30 July 2017).

Method	Converged cases (cloudy 1D-Var)	Converged cases (clear 1D-Var)	Non-converged cases
MR method	69,511 (65%)	29,782 (31%)	8,388 (7%)
ANN method	80,937 (75%)	19,874 (17%)	6,870 (6%)

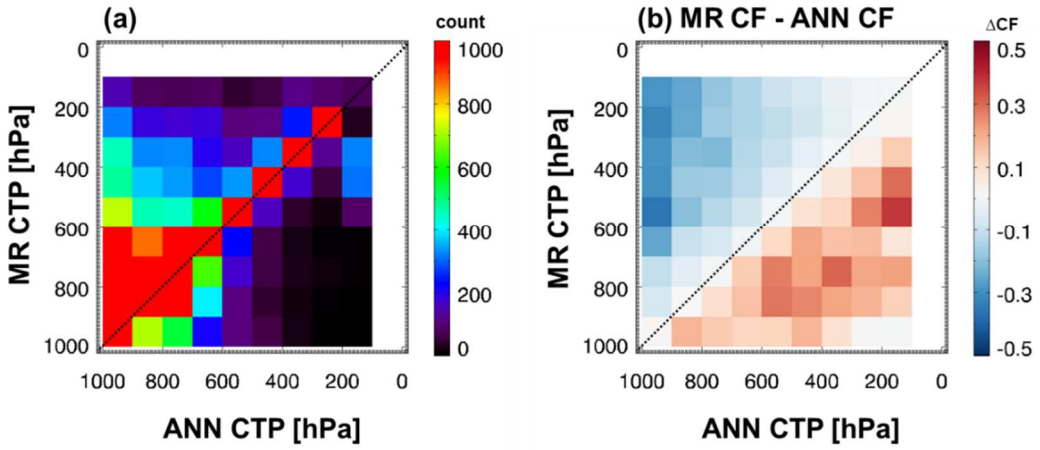


Figure 31. (a) Two-dimensional histograms of frequencies of ANN CTP vs. MR CTP in the UM OPS experiment excluding the scenes having cloud top above 200 hPa level. Color represents data count. (b) Mean difference between MR CF and ANN CF in the same experiment. The mean differences are given in the same CTP pressure coordinates as in the left panels.

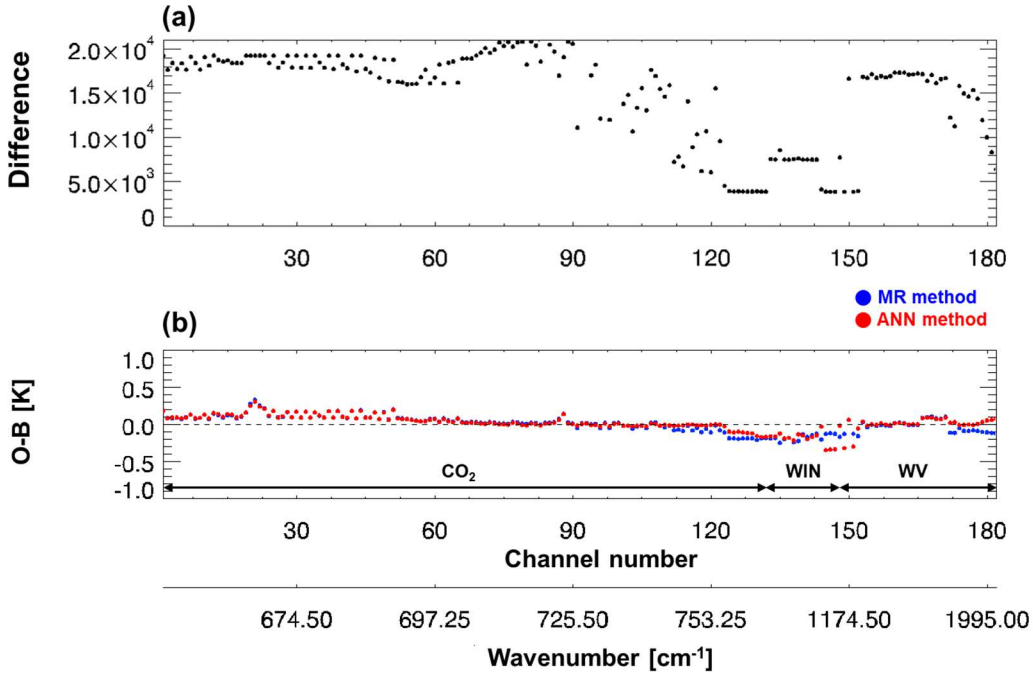


Figure 32. (a) Difference in the number of used measurements (ANN method minus MR method) at each of the 182 IASI channels in the UM OPS experiment excluding the scenes having cloud top 200 hPa level. (b) Mean O-B for the optimum (green), MR method (blue), and for ANN method (red). Black arrows at bottom figure represent IASI channels in CO_2, window, and water vapor absorption bands.

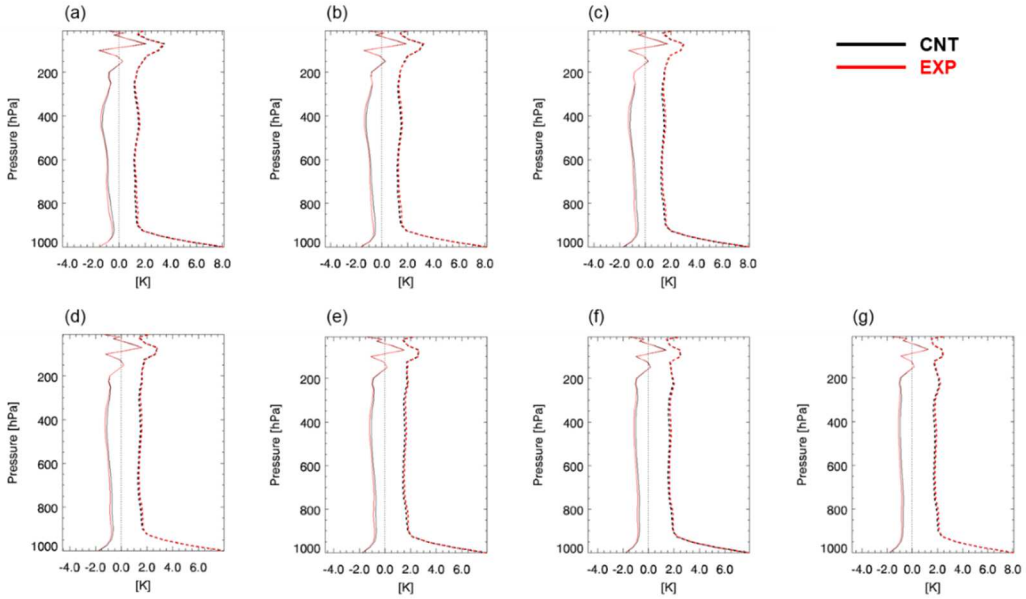


Figure 33. Mean bias (solid lines) and RMSE (dashed lines) profiles of temperature forecasts at (a) T+0, (b) T+12, (c) T+24, (d) T+36, (e) T+48, (f) T+60, and (g) T+72 in the control (black) and the experiment (red) runs.

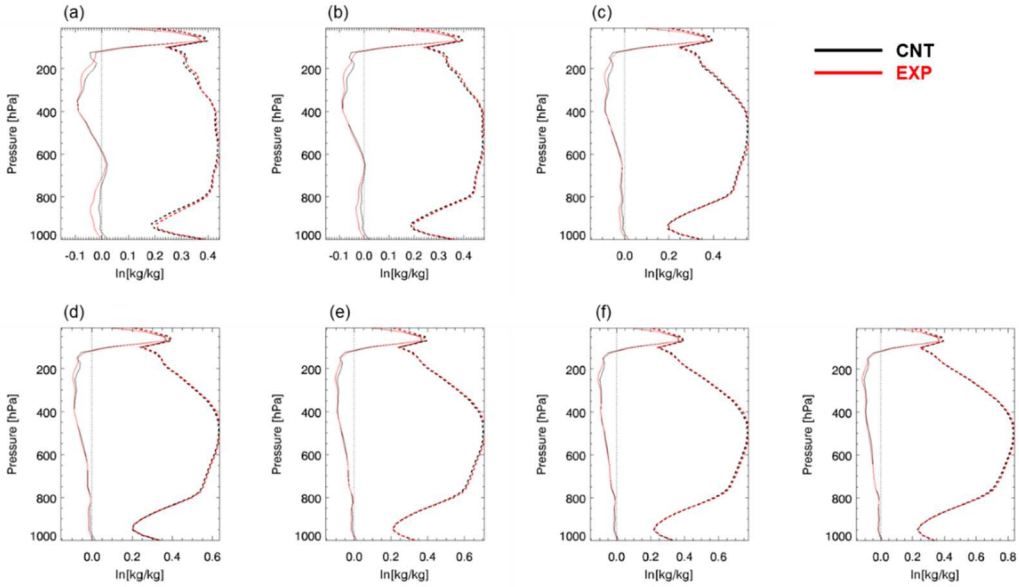


Figure 34. Mean bias (solid lines) and RMSE (dashed lines) profiles of humidity forecasts at (a) T+0, (b) T+12, (c) T+24, (d) T+36, (e) T+48, (f) T+60, and (g) T+72 in the control (black) and the experiment (red) runs.

8.2. Impact on the precipitation forecast

It has been identified that the ANN method increases the use of cloud-affected infrared radiances in the UM IASI data assimilation. However, the increase of used infrared radiances over the cloudy area occupies a very small portion because the use of cloud-affected infrared radiances is very limited compared to the use of clear-sky infrared radiances. This may cause the neutral impact of the ANN method on the UM global temperature and moisture forecasts. Thus, we intend to examine the impact of the ANN method in some specific meteorological phenomena that can be occurred near cloud regions.

The forecast impact of the ANN method on precipitation, which is the most frequent event in the cloudy area, was examined following the methodology in Noh et al (2019). The accuracy of precipitation forecast in the two trial runs was evaluated by use of the Equitable Threat Score (ETS) and Bias Score (BS), and False Alarm Rate (FAR) as verification indices, which are generally used for dichotomous variables (Wilks, 1995).

$$Equitable\ Threat\ Score = \frac{Hit - S_f}{Hit + False\ Alarm + Miss - S_f} \qquad (10)$$

$$S_f = P_c(Hit + False\ Alarm),\quad P_c = \frac{Hit + Miss}{All} \tag{11}$$

$$Bias\ Score = \frac{Hit + False\ Alarm}{Hit + Miss} \tag{12}$$

$$False\ Alarm\ Rate = \frac{False\ Alarm}{False\ Alarm + Correct\ Rejection} \tag{13}$$

'*Hits*' and '*Correct Rejection*' represent the intersection area of precipitation and non-precipitation, respectively, between the model forecast and reference area. '*Misses*' is the precipitation area in the reference data, which was missed in the forecast. '*False Alarms*' indicates the precipitation area in the forecast, which was not assigned as the precipitation area in reference data. The ETS has a value between -1/3 and 1, and a closer to a maximum value of unity represents the higher forecast accuracy. BS larger (smaller) than 1 means that the frequency of the event is overestimated (underestimated). Lastly, the FAR is the ratio of false alarm events, so that the smaller FAR means that the accuracy of the forecast is higher.

As reference data, satellite-based precipitation products from the global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG) were used. The IMERG algorithm is intended to inter-calibrate, merge, and interpolate various satellite microwave precipitation estimates, together with microwave-calibrated infrared satellite estimates, precipitation gauge analyses. In

this study, we used the multi-satellite precipitation product in 12 hours after observation time at a temporal resolution of 30 min and spatial resolution of 0.1°.

Figure 35 shows the verification results of the UM global precipitation forecasts by ETS, BS, and FAR indices by using the IMERG product as a reference. The three indices were used to evaluate all the categories (e.g., hit, miss, false alarm, and correct rejection). The ETS from the experiment run has larger values for all the forecast time range (Figure 35(a)), indicating that the precipitation event is better forecasted in the experiment run. The experiment run has smaller FAR values (Figure 35(b)), representing the less frequency of false-alarm events in the experiment run. It was also identified that both experiments overestimate the precipitation event through the fact that the BS values calculated were larger than unity. The closer the BS value is to unity, the higher the forecast accuracy. It can be seen that the precipitation event is better simulated in the experiment run by having a BS value closer to 1 in the experiment run (Figure 35(c)).

As a result of evaluating the precipitation event with the above three indices, it was confirmed that the accuracy of the UM global precipitation forecasts in the experiment run with the ANN method is high compared to the forecasts in the control run with the MR method. In particular, the spatial distribution of the precipitation forecast has been improved by using the ANN method in the IASI data assimilation system. This may be explained by the fact that the ANN method

uses more IASI radiances in the cloud-sky area, where precipitation event occurs frequently. However, it is still necessary to analyze the variables related to the precipitation event, in order to explain the higher accuracy of the precipitation forecasts, in spite of the neutral impact of the ANN method on the temperature and humidity forecast field.

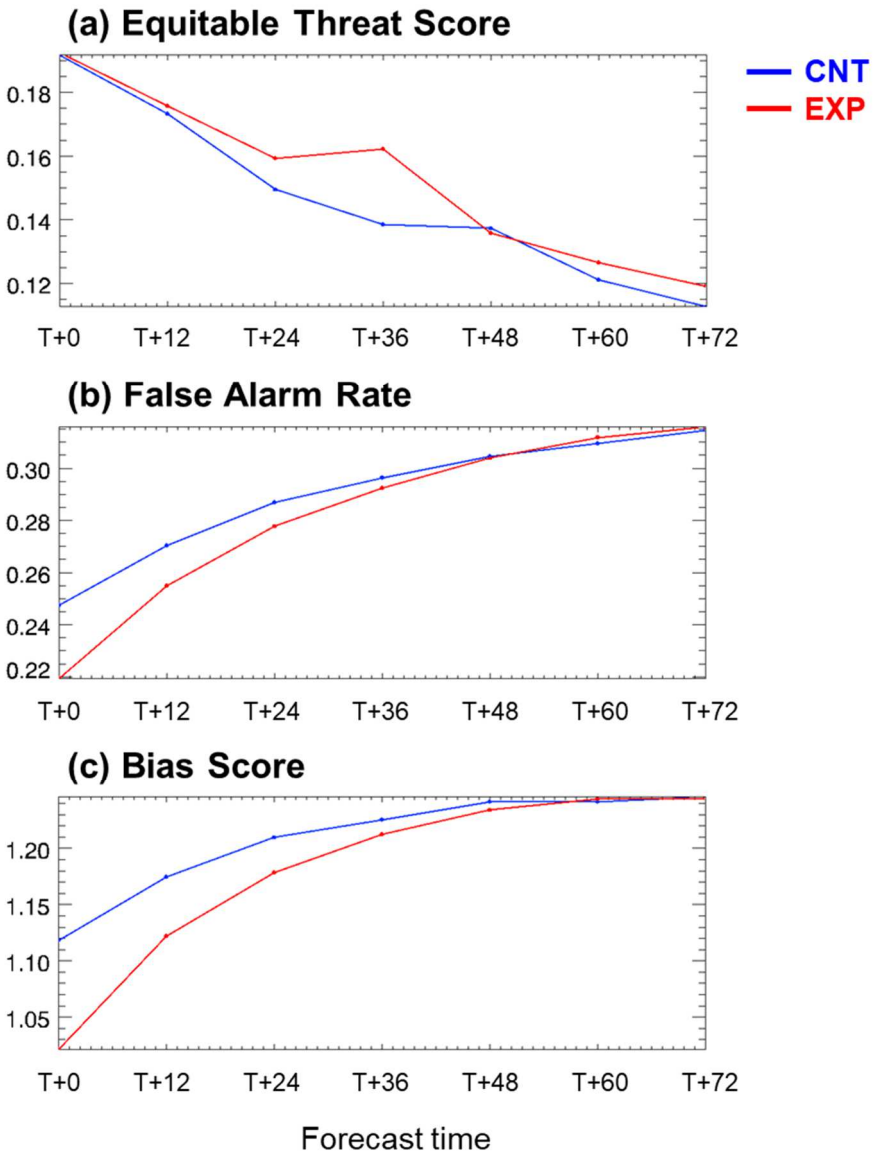


Figure 35. (a) Equitable threat score, (b) false alarm rate, and (c) bias score of global precipitation forecast using GPM IMERG precipitation data as a reference. Blue and red lines indicate the control and experiment runs, respectively.

8.3. Impact of tropical cyclone forecast

In addition to the precipitation events, the forecasts for a tropical cyclone, which is an extreme atmospheric phenomenon that occurs in the cloud region, were evaluated in each experiment. Two tropical cyclone cases (typhoon Noru and hurricane Franklin) during the experiment period were selected and evaluated using the best track data provided by the Joint Typhoon Warning Center (JTWC). The best track from JTWC is estimated using all possible aircraft, satellite, radar, and synoptic data.

Typhoon Noru formed on 19 July 2017 in the Northwest Pacific Ocean and lasted 2 weeks and 6 days. It had a radius of about 350 km and the highest wind speed of about 250 km/h. The tropical cyclone forecasts for the typhoon Noru in the control and experiment runs were assessed. In particular, when it was located near the Korean Peninsula and the Japanese islands, the typhoon track had suddenly changed. Thus, we aim to see whether the forecast for this special event can be improved in the experiment run. Figure 36 shows communication, ocean, and meteorological satellite (COMS) infrared image of typhoon Noru and the difference in the number of used channels in each trial run. There is no difference near the center of the typhoon where the high and thick clouds are dominant. However, in the area surrounding the typhoon, more IASI channels are used in the

experiment run. Thus, it was confirmed that more channels around the typhoon area were used in the IASI data assimilation by using the ANN method.

Figure 37 is a comparison of typhoon Noru tracks in ECMWF, the control run, and the experiment run along with the JTWC best track. The ECMWF best track results are based on the reanalysis data corresponding to the forecast time. In each experimental run, the point of the track was defined as the location with the minimum sea level pressure in the 5°x5° grid box around the JTWC best track. The results from the experiment run show that the typhoon Noru track was forecasted more closely to the JTWC best track for the entire period. The forecasted track in the control run passes the Strait of Korea between the Korean peninsula and the Japanese islands, and it is similar to the forecasted track from the Korea Meteorological Administration. However, in the experiment run, the typhoon track is forecasted to pass over Japan, which was not able to simulate in the control run.

Figure 38(a) and (b) describe the mean RMSE results calculated by comparing the minimum sea level pressure (MSLP) and maximum wind speed (MWS) in the control and experiment runs within 5°x5° grid around the JTWC best track, using the MSLP and MWS data provided by JTWC, respectively. The RMSEs are obtained for the entire period of typhoon Noru. In both variables, it can be seen that the RMSE values of MSLP and MWS in the experiment run are

smaller in the forecast time range from T+0 to T+24. For the forecast time after T+36, the RMSE of MSLP, which is a variable used for the definition of the track, still smaller in the experiment, while the RMSE of MWS has larger errors. Figure 38(c) shows the RMSE of precipitation within 5°x5° grid around the JTWC best track, compared to the GPM IMERG product data. Precipitation forecasts in the experiment run represent an overall neutral impact.

We have confirmed whether the improvement of the track forecast of typhoon Noru was induced by the improved forecast of background states around the typhoon. Figure 39 indicates the temperature, moisture, and geopotential height at 850 hPa level from the 36-hour forecast in each trial compared to the ERA5 reanalysis fields. The forecast results of the temperature, humidity, and geopotential height in the experiment run are closer to the ERA5 reanalysis, which shows that the forecast of the background states near the typhoon is improved. In particular, the result of the 850 hPa geopotential height, which is an important variable for the forecast of typhoon track, shows that the left side of the typhoon has high pressure. This prevents the track of typhoon from going to west, so that the typhoon track is headed to the Japanese island.

The second verification case is hurricane Franklin, which occurred around Veracruz state in Mexico for 3 days from 7 August 2017. Since the diameter of the hurricane was only about 37 km, the point of the track was defined as the location

with the minimum sea level pressure in the 3°x3° grid box around the JTWC best track in this study. Figure 40 describes the Geostationary operational environmental satellite-13 (GOES-13) infrared image and difference in the number of used channels in control and experiment runs. It can be also seen that more channels are used in the experiment run in the regions close to the center of the hurricane.

In Figure 41, the track forecasts of hurricane Franklin, predicted at 12 UTC 06 August 2017 and 00 UTC 07 August 2017 before it reached over land, are closer to the JTWC best track in the experiment run. Additionally, when the RMSEs of MSLP, MWS, and precipitation in the 3°x3° grid box around the JTWC best track are compared in both trial runs (Figure 42), the RMSEs at T+12 and T+24 appeared to be smaller in the experiment run. Figure 43 shows the difference between ERA5 reanalysis and 48-hour forecast of 850 hPa geopotential height in the control run and how this difference was changed in the experiment run. The result represents that the difference from ERA5 reanalysis shown in the control run is reduced in the experiment run. The lower pressure near the central part of the hurricane in the experiment run makes the forecast track closer to the best track. Considering that the level of confidence in the forecast field for such a short-lasting case is less reliable, it is considered that better hurricane simulation was performed in the experiment run.

It was demonstrated that it is possible to simulate and forecast extreme weather events occurring in a short period of time by the implementation of the ANN method in the IASI data assimilation process in the UM OPS. The ANN method was examined to have a positive impact on not only the large system such as typhoon Noru but also on the relatively small tropical cyclone. In particular, the impact of the ANN method on the forecasts was noticeable in these case studies because meteorological phenomena such as tropical cyclone exist over a wide area of clouds. As a result, it would be necessary to clearly see the forecast impact on the UM forecast field by using more IASI radiance data from the ANN method in cloud regions.

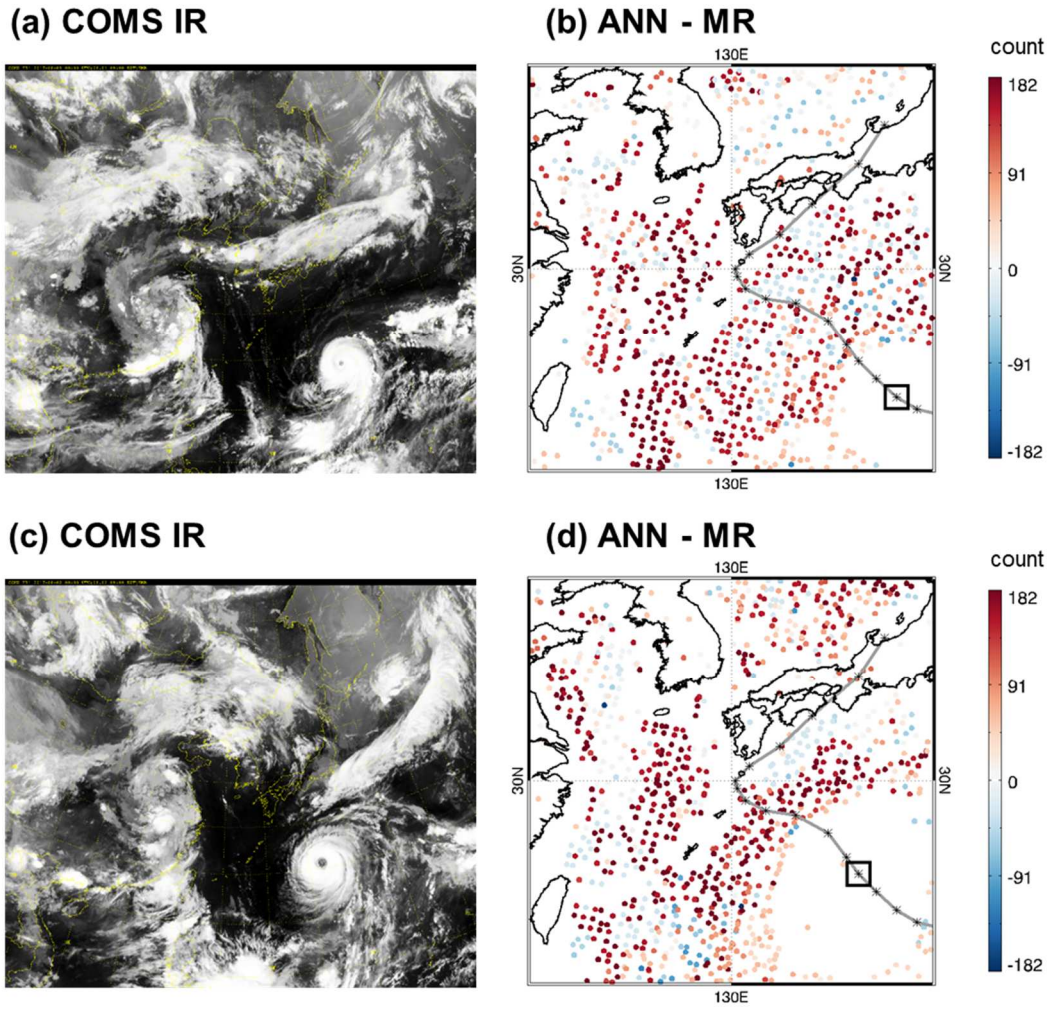


Figure 36. (a) COMS 10.8 ㎛ image and (b) difference in the number of used IASI channels between in the control run and in the experiment run (experiment minus control) at 00 UTC 01 August. (c) and (d) are at 00 UTC 02 August. Grey line represents the JTWC best track and the black square shows the center of the typhoon Noru.

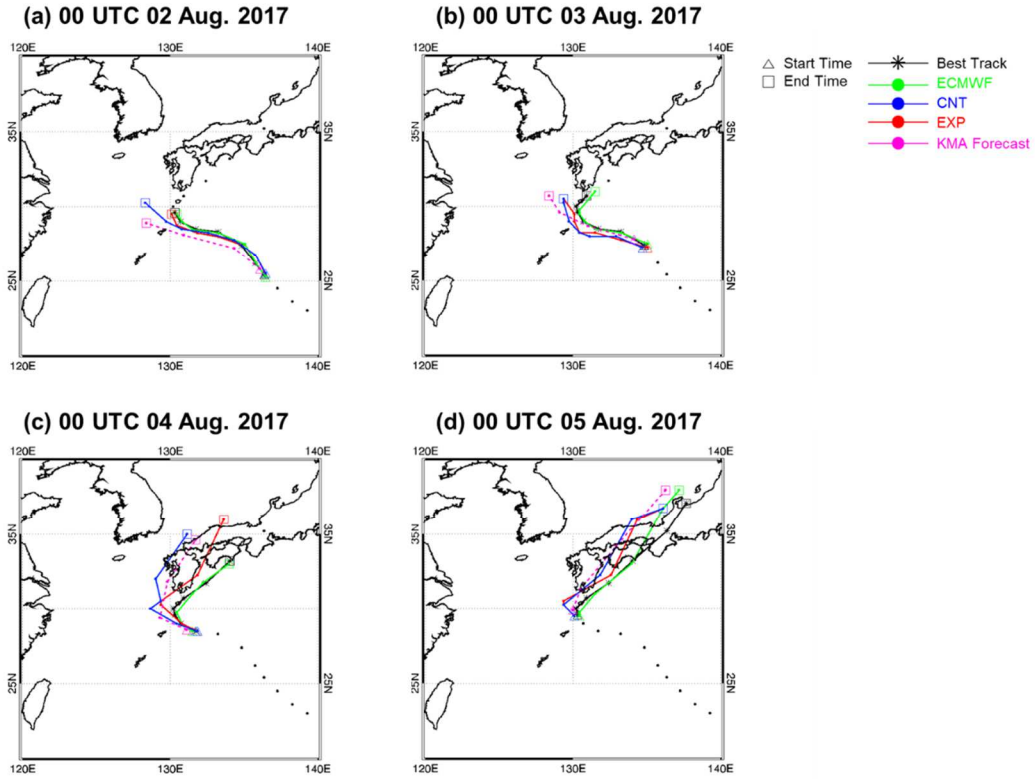


Figure 37. Typhoon Noru track of the JTWC best track (black with asterisk), the ECMWF reanalysis (green), the KMA forecast (purple), and forecasts from (a) 00 UTC 02 August, (b) 00 UTC 03 August, (c) 00 UTC 04 August, and (d) 00 UTC 05 August 2017 in the control (blue) and the experiment (red) runs. Black dots show every 12-hour track of the whole period of typhoon Noru and triangle and square represent the track at T+0 and T+72 forecast, respectively.

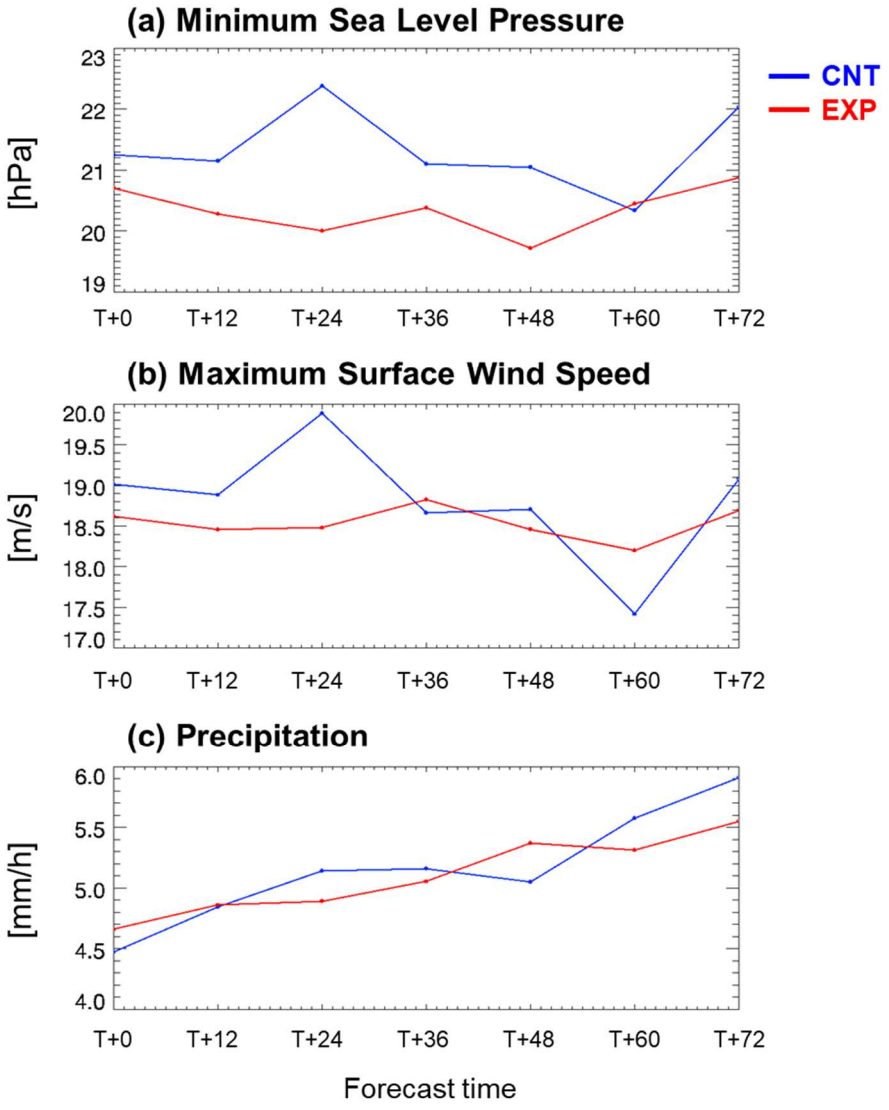


Figure 38. RMSE of (a) minimum sea level pressure and (b) maximum surface wind speed at T+0 to T+72 forecasts for the typhoon Noru in the control (blue) and experiment (red) runs using the JTWC best track data as a reference. (c) RMSE of precipitation forecasts in both runs is calculated against the GPM IMERG data. Blue (red) represents the results in the control (experiment) run.

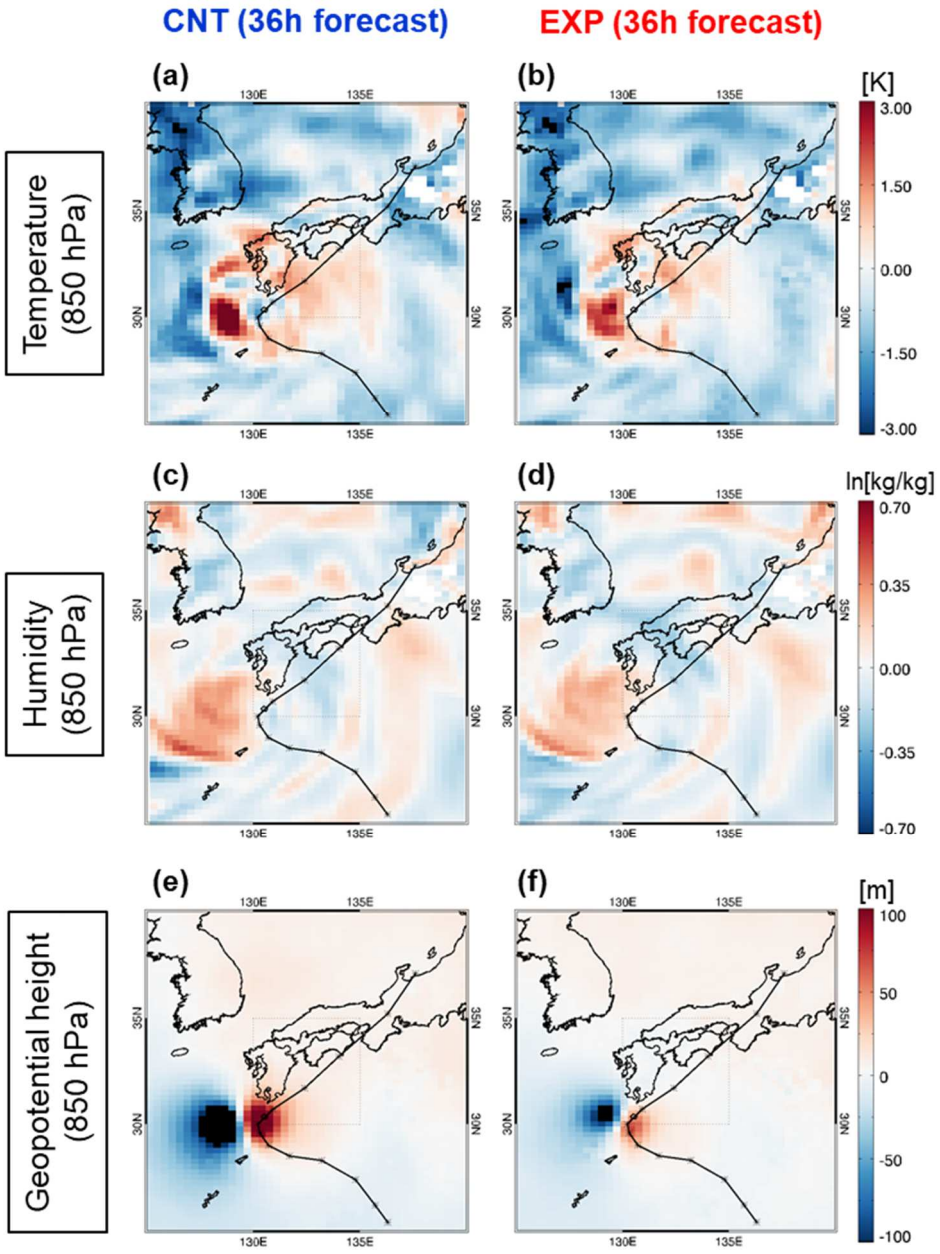


Figure 39. Difference between the ERA5 reanalysis and the 36-hour forecast of (a), (b) 850 hPa temperature, (c), (d) 850 hPa humidity, and (e), (f) 850 hPa geopotential height from 00 UTC 04 August in the control (left) and the experiment (right) runs, respectively. The difference is the 36-hour forecast minus ERA5 reanalysis and black line represents the JTWC best track.

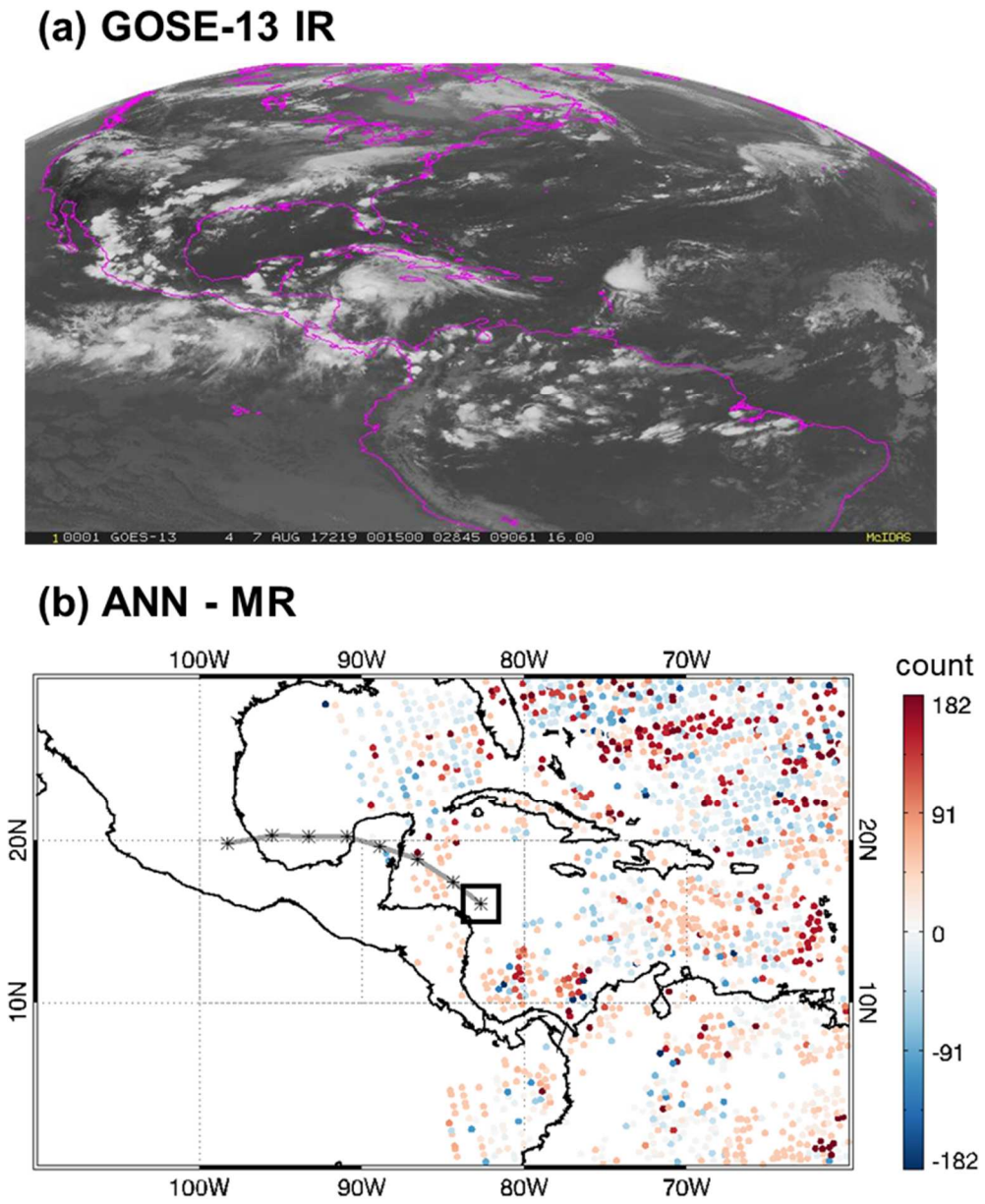


Figure 40. (a) GOES-13 10.7 μm image and (b) difference in the number of used IASI channels between in the control run and in the experiment run (experiment minus control) at 00 UTC 07 August. Grey line represents the JTWC best track and the black square shows the center of the hurricane Franklin.

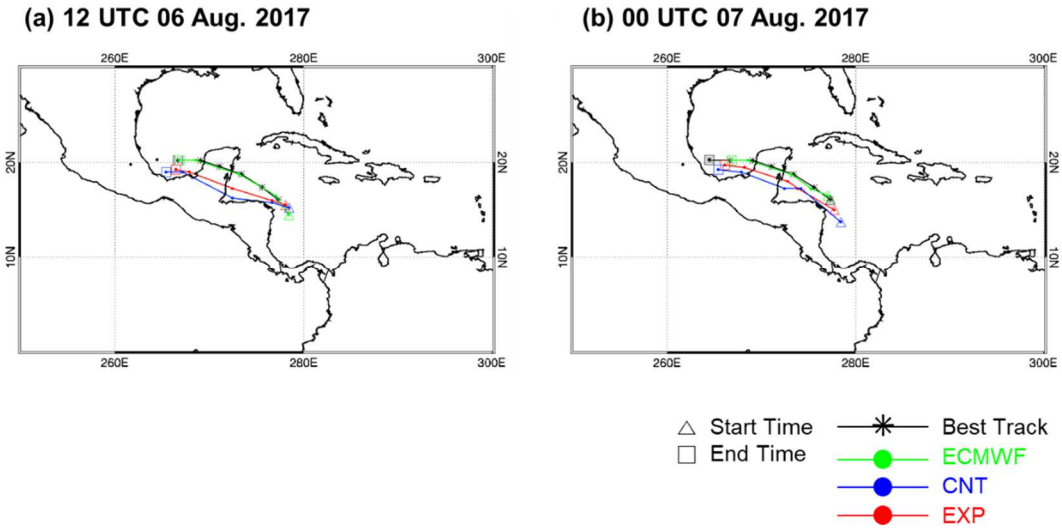


Figure 41. Hurricane Franklin track of the best track (black with asterisk), the ECMWF reanalysis (green), and 72h forecasts from (a) 12 UTC 06 August, (b) 00 UTC 07 August 2017 in the control (blue) and the experiment (red) runs. Black dots show every 12-hour track of the whole period of hurricane Franklin and triangle and square represent the best track at T+0 and T+72 forecast, respectively.

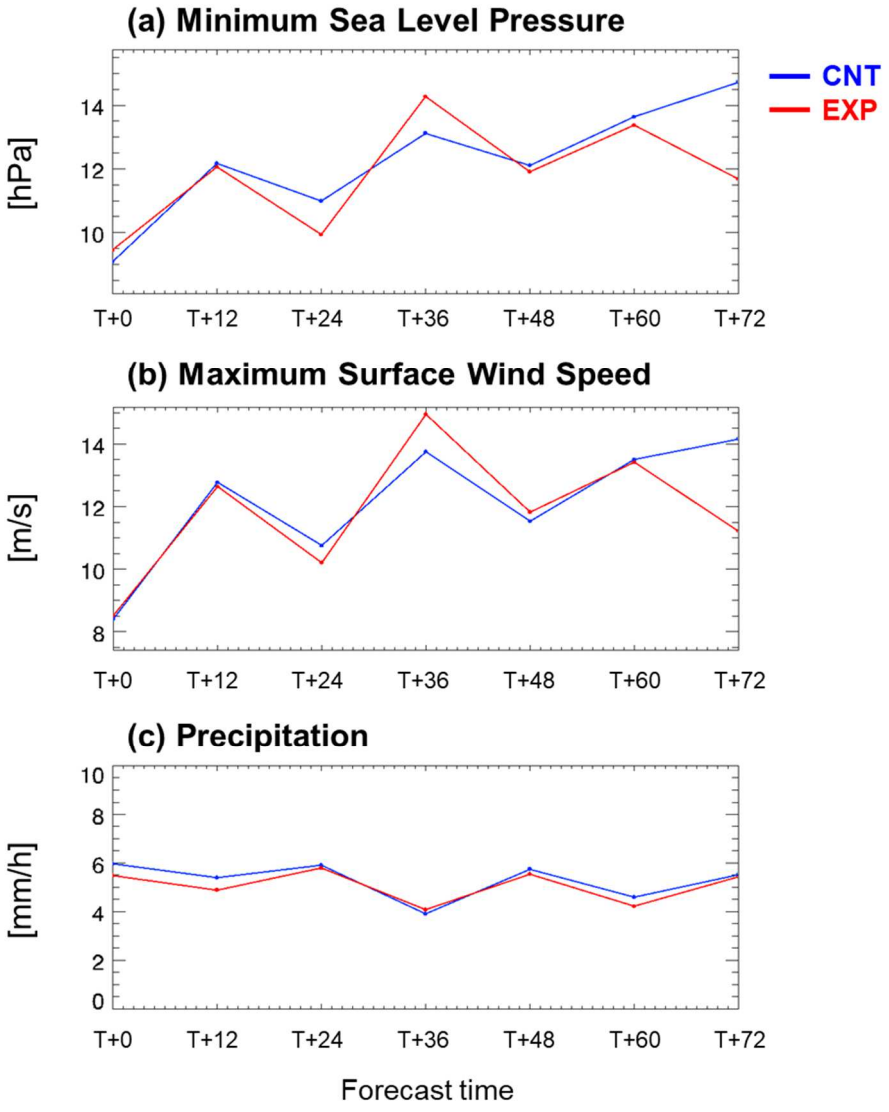


Figure 42. RMSE of (a) minimum sea level pressure and (b) maximum surface wind speed forecast at T+0 to T+72 for the hurricane Franklin in the control (blue) and experiment (red) runs using JTWC best track data as a reference. (c) RMSE of precipitation forecasts in both runs is calculated against the GPM IMERG data. Blue (red) represents the results in the control (experiment) run.

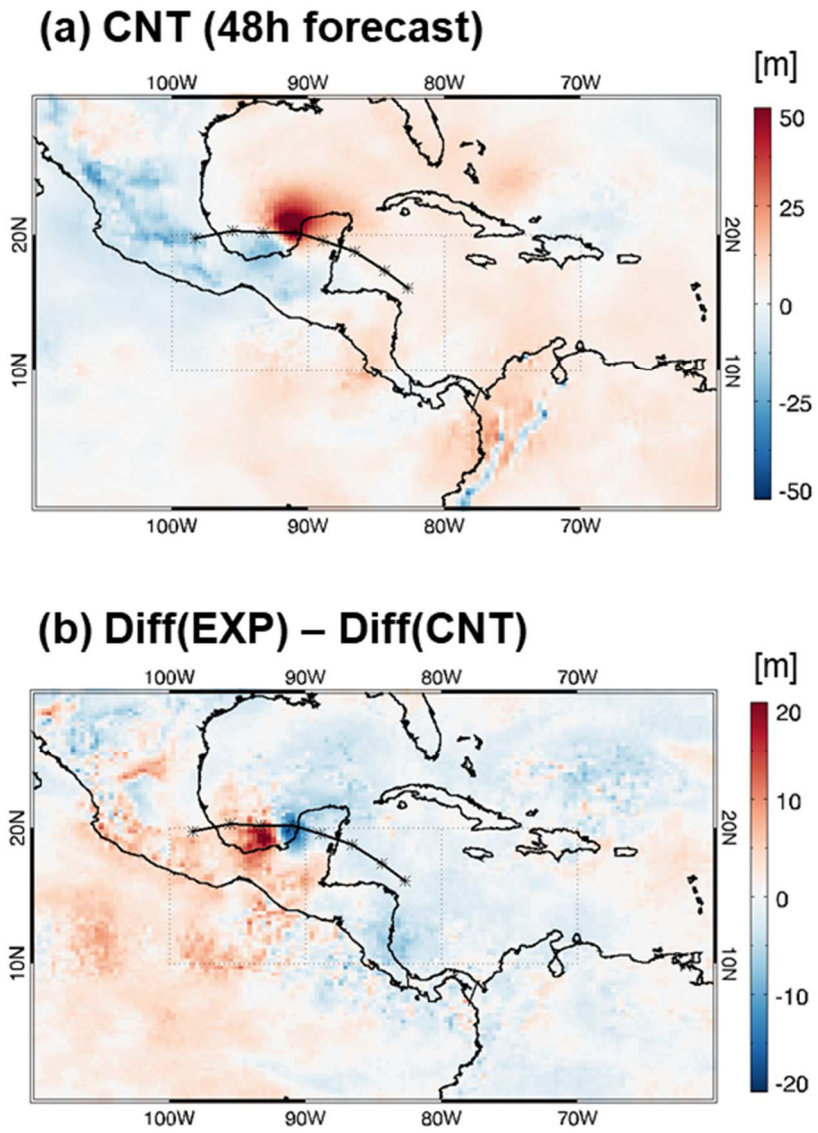


Figure 43. (a) Difference between the ERA5 reanalysis and the 48-hour forecast of 850 hPa geopotential height from 00 UTC 07 August in the control. (b) The difference between the ERA5 reanalysis and the 48-hour forecast of 850 hPa geopotential height from 00 UTC 07 August in the experiment run minus the difference in the control run. The difference means the 48-hour forecast minus ERA5 reanalysis and black line represents the JTWC best track.

9. Summary and discussion

There have been efforts of directly using cloud-affected (or not clear) hyperspectral infrared radiances, instead of indirect use like removing cloud-contaminated channels or constructing ‘presumed’ clear-sky radiances. In the Cloudy 1D-Var method employed by the IASI 1D-Var data assimilation system, the first guess pair of CTP and CF is assigned from infrared radiances of 10 IASI channels and model background field by minimizing the residual between observed and simulated channel radiances. In this approach, clouds are assumed to be a single layer behaving like a grey body. Although the Cloudy 1D-Var method led to a better analysis performance, compared to the clear-sky only 1D-Var or McNally and Watts (2003) method, it is difficult to evaluate the accuracy of obtained cloud properties or examine how we improve the cloud retrieval capability in the Cloudy 1D-Var method.

In this study, we evaluated the cloud retrieval in the Cloudy 1D-Var method (i.e. MR method). Instead of comparing retrieved cloud properties with observations, retrievals were compared against the optimum cloud parameters which yield the best 1D-Var analysis results, given IASI observations and model background field. For the comparison, atmospheric temperature and humidity profiles as well as cloud information from the ECMWF short-range forecasts were

considered to be the truth, from which IASI radiances and background fields were generated after taking their respective errors into account. The comparison indicated that cloud retrieval from the MR method tends to overestimate the cloud top height and thus tends to underestimate the cloud fraction. The overestimation brought in less active use of cloud-affected radiances, necessitating the improvement of the cloud retrieval method by correcting this overestimation of the cloud top height.

In order to improve the cloud retrieval, we developed a method of resembling the optimum cloud parameters as closely as possible. In doing so, we utilized an artificial neural network approach to train the inputs (IASI 182 channel radiances and model background fields) to produce cloud parameters similar to the optimum values. It was shown that the ANN method produces cloud tops closer to the optimum values, compared to the results from the MR method. It was also found that corresponding CFs are in better agreement with optimum values. Moreover, it was noted that the ANN method gave more converged cases in the 1D-Var analysis. It is believed that the ANN method allows the use of more cloud-affected channels, with similar O-Bs, and reduced RMSEs in the 1D-Var analysis.

The impact of the ANN method on the 1D-Var analysis was examined within the UM OPS by taking experiments over 30 days (from 15 July to 13 August 2017). In these experiments, retrieved cloud tops from the ANN method were generally

lower than in the control run with the MR method. Additionally, more CO_2 channels were selected, with O-B means nearly the same as found from the MR method. In spite of less channels used by the ANN method in the UM OPS, the overall convergence cases by the ANN method was 1% more than by the MR method. Furthermore, analysis results for the ANN method at T+0 forecast time is found to be neutral. Thus, the use of more channels (mostly window channels) by the MR method may not be always beneficial if more converged cases and the neutral analysis impact by the ANN method are considered. Given that the ANN method gives more converged cloudy cases, it tends to compensate less window channels with more CO_2 and water vapor channels.

Another experiment was conducted to assess the impact of the ANN method on the UM forecast by rejecting the cases having higher cloud above the 200 hPa level in both trial runs. The features of making the retrieved CTP lower and using more channel information were also shown in this experiment, resulting rather neutral impact on the UM T+0 to T+72 forecasts. Nevertheless, the forecast results of some atmospheric phenomena such as precipitation and tropical cyclone near cloudy areas show less forecast error in the experiment run. This suggests that the implementation of the ANN method in the UM OPS could improve the forecasts of the dynamic atmospheric events occurring near cloudy areas. As far as the impact of the ANN method on the global forecasts of temperature and moisture

field is neutral, further study into which mechanisms have improved the precipitation forecast and which atmospheric states have affected the forecasts for the extreme weather event will be encouraging.

Even if the impact is near neutral at the T+0 analysis in the UM OPS, the implementation of the ANN method should be beneficial because the ANN method can substantially reduce the computational burden to search for an initial guess pair of CTP and CF. To find a CTP and CF pair showing a minimum residual, the MR method needs to calculate at least 27 CTP and CF pairs (from 100 hPa level to the surface level) and corresponding residuals. By contrast, in the ANN method, CTP is directly retrieved with ANN-derived coefficients, and then CF is determined using Eq. (6). Therefore, if the ANN method is introduced to the IASI 1D-Var data assimilation system, the computational time required for finding initial cloud parameters can be reduced to roughly 1/54 level (approximately 1.85%) of what required for the MR method. Furthermore, the ANN method may avoid problems of possible multiple solutions in the MR method because the CTP retrieval is independent of CF in the ANN method.

On the other hand, the impact of the ANN method on the 1D-Var analysis appears neutral; analysis results are not much discernable from those by the MR method. It is likely caused by sub-optimal observation and background error covariance matrices (i.e. **R** matrix and **B** matrix) employed in the UM OPS which

represent the clear-sky field of view. As shown in Figure 2 and Figure 3, both cloud retrieval methods find retrieved cloud parameters likely located along the line showing approximately an inverse relationship between CTP and CF. Concerning the MR method, which tends to overestimate the cloud height, retrieved CF should be smaller than optimum values. Thus, the MR method will likely identify the cloud-affected scenes more often to be clear because the criteria determining cloud presence is CF > 0.05. By contrast, the ANN method will more likely identify the same cloud-affected scenes to be cloudy, which are closer to optimum values. Since the ANN method allows more cloud-affected IASI measurements, as seen in the results, there will be more added CO_2 channels. However, low-level cloud-affected channels by the ANN method seem to be rejected during the data assimilation process. Note that observation and radiative simulation of channel radiances are subject to larger errors for cloud-sky cases, in comparison to the clear-sky cases. In particular, it is generally known that the radiative transfer simulation for the cloud-sky scenes creates much larger error than the clear-sky simulation. These can cause less use of lower-level peaking channels and window channels, compared to MR method results. Thus, even if the ANN method employs more cloud-affected scenes, those may not be fully utilized in the assimilation, giving near neutral results probably by adding CO_2 channels but losing window channels. Thus, one way to fully accommodate the ANN-retrieved

CTP and CF in the UM OPS may be either to use observation and background error covariance matrices for the cloudy-sky scene or to allow larger uncertainty errors for the cloud-affected channels in the UM OPS. Further studies should be done along the line of those directions, to actively use cloud-affected IASI scenes in the data assimilation.

References

Amorati, R., and R. Rizzi, 2002: Radiances simulated in the presence of clouds by use of a fast radiative transfer model and a multiple-scattering scheme. *Appl. Opt.*, **41**, 1604-1614, https://doi.org/10.1364/AO.41.001604

August, T., et al., 2012: IASI on Metop-A: Operational Level 2 retrievals after five years in orbit. *J. Quant. Spectrosc. Radiat. Transf.*, **113**, 1340-1371, https://doi.org/10.1016/j.jqsrt.2012.02.028

Auligné, T., A. P. McNally, and D. P. Dee, 2007: Adaptive bias correction for satellite data in a numerical weather prediction system. *Q. J. R. Meteorol. Soc.*, **133**, 631-642, https://doi.org/10.1002/qj.56

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47-55, https://doi.org/10.1038/nature14956

Boudala, F. S., G. A. Isaac, Q. Fu, and S. G. Cober, 2002: Parameterization of effective ice particle size for high-latitude clouds. *Int. J. Climatol*, **22**, 1267-1284, https://doi.org/10.1002/joc.774

Collard, A. D., 2007: Selection of IASI channels for use in numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **133**, 1977-1991, https://doi.org/10.1002/qj.178

Chevallier, F., S. Di Michele, and A. P. McNally, 2006: Diverse profile datasets from the ECMWF 91-level short-range forecasts. ECMWF, reading, United Kingdom.

Desai, K. M., S. A. Survase, P. S. Saudagar, S. S. Lele, and R. S. Singhal, 2008:

Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: case study of fermentative production of scleroglucan. *Biochem. Eng. J.*, **41**, 266-273, https://doi.org/10.1016/j.bej.2008.05.009

Diebel, D., F. Cayla, T. Phulpin, P. Courtier, and M. Langevin, 1996: IASI mission rationale, and requirements. *Tech. Rep. IA-SM-0000–10-CNE/EUM*, EUMETSAT, Darmstadt, Germany.

Eresmaa, R., and A. P. McNally, 2014: Diverse profile datasets from the ECMWF 137-level short-range forecasts. NWP SAF Report No. NWPSAF-EC-TR-017

Eyre, J. R., and W. P. Menzel, 1989: Retrieval of cloud parameters from satellite sounder data: A simulation study. *J. Appl. Meteorol.*, **28**, 267-275, https://doi.org/10.1175/1520-0450(1989)028<0267:ROCPFS>2.0.CO;2

Faijan, F., L. Lavanant, and F. Rabier, 2012: Towards the use of cloud microphysical properties to simulate IASI spectra in an operational context. *J. Geophys. Res.*, **117**, D22205, https://doi.org/10.1029/2012JD017962

Hilton, F., N. C. Atkinson, S. J. English, and J. R. Eyre, 2009: Assimilation of IASI at the Met Office and assessment of its impact through observing system experiments. *Q. J. R. Meteorol. Soc.*, **135**, 495–505, https://doi.org/10.1002/qj.379

Hilton, F., et al., 2012: Hyperspectral Earth observation from IASI: Five years of accomplishments. *Bull. Am. Meteorol. Soc.*, **93**, 347-370, https://doi.org/10.1175/BAMS-D-11-00027.1

Joo, S., J. R. Eyre, and R. Marriott, 2013: The impact of Metop and other satellite data within the Met Office global NWP system using an adjoint-based sensitivity method. *Mon. Weather Rev.*, **141**, 3331-3342, https://doi.org/10.1175/MWR-D-12-00232.1

Klaes, K. D., et al., 2007: An introduction to the EUMETSAT polar system. *Bull. Am. Meteorol. Soc.*, **88**, 1085-1096, https://doi.org/10.1175/BAMS-88-7-1085

Le Marshall, J., et al., 2006: Improving global analysis and forecasting with AIRS. *Bull. Am. Meteorol. Soc.*, **87**, 891-894.

Lerner, J. A., E. Weisz, and G. Kirchengast, 2002: Temperature and humidity retrieval from simulated Infrared Atmospheric Sounding Interferometer (IASI) measurements. *J. Geophys. Res. Atmos.*, **107**, D14, https://doi.org /10.1029/2001JD900254.

Li, J., et al., 2005: Optimal cloud-clearing for AIRS radiances using MODIS. *IEEE Trans. Geosci. Remote Sens.*, **43**, 1266-1278, https://doi.org /10.1109/TGRS.2005.847795

Martinet, P., et al., 2014: Evaluation of a revised IASI channel selection for cloudy retrievals with a focus on the Mediterranean basin. *Q. J. R. Meteorol. Soc.*, **140**, 1563-1577, https://doi.org/10.1002/qj.2239

McFarquhar, G. M., S. Iacobellis, and R. C. Somerville, 2003: SCM simulations of tropical ice clouds using observationally based parameterizations of microphysics. *J. Clim.*, **16**, 1643-1664, https://doi.org/10.1175/1520-0442(2003)016<1643:SSOTIC>2.0.CO;2

McNally, A. P. and P. D. Watts, 2003: A cloud detection algorithm for high-spectral-resolution infrared sounders. *Q. J. R. Meteorol. Soc.*, **129**, 3411–3423, https://doi.org/10.1256/qj.02.208

McNally, A. P., et al., 2006: The assimilation of AIRS radiance data at ECMWF. *Q. J. R. Meteorol. Soc.*, **132**, 935–957, https://doi.org/10.1256/qj.04.171

McNally, T., M. Bonavita, and J. N. Thépaut, 2014: The role of satellite data in the forecasting of Hurricane Sandy. *Mon. Weather Rev.*, **142**, 634-646, https://doi.org/10.1175/MWR-D-13-00170.1

Migliorini, S., 2015: Optimal ensemble-based selection of channels from advanced sounders in the presence of cloud. *Mon. Weather Rev.*, **143**, 3754-3773, https://doi.org/10.1175/MWR-D-14-00249.1

Noh, Y. C., B. J. Sohn, and Y. Kim, 2020: The Impact of a New Set of IASI Channels on the Unified Model Global Precipitation Forecast. *Asia-Pac. J. Atmos. Sci.*, **56**, 45-56, https://doi.org/10.1007/s13143-019-00129-6

Okamoto, K., A. P. McNally, and W. Bell, 2014: Progress towards the assimilation of all-sky infrared radiances: an evaluation of cloud effects. *Q. J. R. Meteorol. Soc.*, **140**, 1603-1614, https://doi.org/10.1002/qj.2242

Ou, S. C., and K. N. Liou, 1995: Ice microphysics and climatic temperature feedback. *Atmos. Res.*, **35**, 127-138, https://doi.org/10.1016/0169-8095(94)00014-5

Pavelin, E. G., S. J. English, and J. R. Eyre, 2008: The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **134**, 737–749, https://doi.org/10.1002/qj.243

Prunet, P., J.-N. Thépaut, and V. Cassé, 1998: The information content of clear sky IASI radiances and their potential for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **124**, 211–241, https://doi.org/10.1002/qj.49712454510

Rawlins, F., et al., 2007: The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **133,** 347–362, https://doi.org/10.1002/qj.32

Rodgers C. D., 2000: Inverse methods for atmospheric sounding: theory and practice. World Scientific: Singapore.

Rumelhart, D. E., G. E. Hinton, and J. L. McClelland, 1986: A general framework for parallel distributed processing. Parallel distributed processing: Explorations in the microstructure of cognition, **1**, 45-76.

Smith, W. L., et al., 2009: Evolution, current capabilities, and future advance in satellite nadir viewing ultra-spectral IR sounding of the lower atmosphere. *Atmos. Chem. Phys.*, **9**.

Wilks, D. S., 2011: Statistical methods in the atmospheric sciences (Vol. 100). *Academic press*, New York.

Wylie, D. P., and W. P. Menzel, 1999: Eight years of high cloud statistics using HIRS. *J. Clim.*, **12**, 170-184, https://doi.org/10.1175/1520-0442(1999)012<0170:EYOHCS>2.0.CO;2

Wyser, K., 1998: The effective radius in ice clouds. *J. Clim.*, **11**, 1793-1802, https://doi.org/10.1175/1520-0442(1998)011<1793:TERIIC>2.0.CO;2

국문초록

기상청 현업 모델인 통합수치모델 (Unified Model) 내 자료동화 과정 중, 적외 초분광 센서인 IASI (Infrared Atmospheric Sounding Interferometer) 관측자료를 활용하는 방법으로는 구름 변수를 1D-Var 과정 내에서 산출하는 Cloudy 1D-Var 방법(Pavelin et al., 2008)이 사용되고 있다. Cloudy 1D-Var 방법에서는 산출된 구름변수(운정고도, 운량)를 이용해 구름지역을 탐지할 뿐만 아니라 자료동화에 사용되는 채널을 선정하기 때문에 운정고도와 운량을 정확하게 산출하는 것은 매우 중요하다. Cloudy 1D-Var 방법에서 구름 변수인 운정고도와 운량의 초기값은 minimum residual (MR) 방법(Eyre and Menzel, 1989)을 통해 관측 복사량과 모델 배경장이 만들어내는 복사량의 차이를 최소화 시키는 값으로 얻어진다. 본 연구에서는 ECMWF 단기예보장을 활용하여 IASI 모의 관측 복사량과 모델 배경장을 생산하였고, 이를 이용해 최종 온도 습도 분석장의 에러를 최소로 만드는 새로운 구름변수들을 찾아내어 이를 최적의 운정고도, 운량으로 정의하였다. 정의한 최적의 구름 변수를 기준으로 MR 방법이 구름의 고도를 상대적으로 상층으로 산출한다는 것을 확인하였고, 이로 인해 온습도 1D-Var 분석장의 에러를 최소화 시키지 못한다는 점을 확인하였다.

온습도 1D-Var 분석장을 개선하기 위해 최적의 구름변수와 가장 가까운 구름변수를 산출할 수 있는 방법을 모색하였고, IASI 적외 관측 복사량과 모델 배경장을 입력 자료로 하여 운정고도를 산출해내는 인공신경망(ANN;

Artificial Neural Network) 모델을 개발하였다. 검증을 통해 ANN 모델에서 산출된 운정고도, 운량이 앞서 정의한 최적의 운정고도, 운량과 높은 상관관계를 갖는다는 것을 확인하였고, 구름에 의해 영향을 받은 더 많은 채널들이 자료동화 과정 내에 사용되는 것을 확인하였다. 이와 함께 기존 MR 방법을 사용했을 때 얻어진 1D-Var 온습도 분석장 결과와 비교해보았을 때 모든 층에서 온습도 분석장이 개선되는 것을 볼 수 있었고, 특히 중층에서 온도 에러가 10% 가량 줄어드는 것을 확인하였다. 개발한 ANN 모델을 이용하면 운정고도를 먼저 산출하고, 이를 이용해 운량을 계산한다는 점에서 계산시간을 기존 MR 방법의 1.85%로 줄이는 장점까지 얻을 수 있었다.

또한 새로 개발한 ANN 알고리즘을 실제 UM 내에도 적용시켜 보았는데, 이때도 새롭게 산출된 운정고도가 기존의 MR 방법을 통해 산출되었던 운정고도보다 상대적으로 낮게 산출되면서 더 많은 구름지역 IASI 적외 초분광 채널 정보가 자료동화 과정 내에 사용된다는 것을 확인할 수 있었다. 나아가 새롭게 개발한 ANN 방법이 수치예보 모델 초기장 및 예보장 정확도에 주는 영향도 살펴보았다. 전 지구적 온습도 초기장 및 예보 정확도에 미치는 영향은 미미하게 나타났지만, 주로 구름 지역 주변에서 구름을 동반하여 발생하는 날씨 현상인 강수 및 열대 저기압의 예보정확도가 ANN 방법을 사용함으로써 향상되는 것을 확인할 수 있었다.

주요어: 적외 초분광 관측자료, IASI, 자료동화, 구름지역 1D-Var 분석장, 인공신경망 방법, 수치예보

학 번: 2012-20346

감사의 글

박사학위논문을 마무리하는 지금, 정신적으로 힘들 때나 연구가 잘 풀리지 않아 힘들 때 옆에서 힘이 되어 주신 분들이 참 많이 떠오르는 것을 보면 저는 참 인복이 많은 사람인 것 같습니다. 앞으로 살면서 하나하나 갚아 나갈 수 있는 박사가 되겠습니다.

턱없이 부족했던 제가 한 발짝씩 앞으로 나아가 여기까지 올 수 있게 이끌어 주신 손병주 교수님께 가장 먼저 감사의 인사를 드립니다. 연구에 사소한 부분은 없다며 가르쳐 주신, 박사로서 가져야 할 적극적인 연구 태도를 앞으로도 잊지 않겠습니다. 바쁘신 와중에도 학위 위원을 흔쾌히 맡아 주시고, 학위 논문을 완성할 수 있게 적극적으로 도와주신 위원분들께도 감사드립니다. 제 연구의 장점을 부각시켜 주시고 격려해주신 위원장 손석우 교수님, 연구 수행이 힘들 때 흔쾌히 도움 주시고 항상 친근하게 챙겨 주시는 안명환 교수님, 저의 박사 연구 생활 롤모델이자 저의 영원한 김박사님 김윤재 센터장님, 제 연구를 적극적으로 함께 고민해주시고 항상 응원해주시는 전형욱 박사님 정말 감사합니다. 또한 엄청난 연륜에도 학문적, 연구적으로 항상 동료처럼 대해 주시는 영국 기상청 Roger Saunders 박사님과 영국 출장 기간 동안 많은 것을 알려주시고 계속 옆에서 챙겨 주신 Ed Pavelin 박사님께도 먼 한국에서 감사의 마음을 전합니다.

대기과학이라는 학문에 재미를 느낄 수 있게 가르쳐 주신 이동규 교수님,

전종갑 교수님, 윤순창 교수님, 강인식 교수님, 임규호 교수님, 김광열 교수님, 최우갑 교수님, 백종진 교수님, 허창회 교수님, 박록진 교수님, 김상우 교수님께도 감사드립니다. 그리고 대학원 생활에서 소홀하게 지나갈 수 있는 부분들이 흩어지지 않게 챙겨 주신 류현희 조교님, 허종윤 조교님, 김종원 조교님께도 감사드립니다.

함께 실험실 생활을 하진 않았지만, 항상 귀엽게 봐주시고 만날 때마다 따뜻하게 조언과 위로까지 해주셨던 실험실 대선배님들, 신도식 선배님, 성찬오빠, 도형오빠, 근혁오빠, 현종오빠, 상삼오빠, 은희언니, 아영언니 매번 감사합니다. 같이 대학원 생활을 할 때부터 지금까지 항상 제 고민 들어주시고 따뜻하게 응원해주시는 효진언니, 대학원 생활이 제게 준 가장 큰 선물인 친오빠 같은 두 오빠, 센스쟁이 현성오빠랑 잔소리쟁이 영찬오빠가 없었으면 절대 여기까지 오지 못했을 겁니다. 정말 고맙습니다. 그리고 쓴소리 대왕이지만 속은 따뜻한 환진오빠, 대학원 생활을 즐겁게 시작할 수 있게 함께 해주신 재창오빠, 민진언니 부부, 같은 실험실 생활을 하면서 대학교 때보다 친해진 친구이자 고마운 선배님 상무, 만날 때마다 즐거운 센스쟁이 병권이, 타국에서 힘들게 공부하면서도 오히려 저를 응원해줬던 데아, 지금은 인생 선배님이 돼서 예쁜 말로 뭐든 먼저 챙겨주는 송희 모두 감사합니다. 제가 가장 예민한 시기에 함께 실험실 생활하느라 눈치 보며 고생했을 실원들, 안 맞는 것 같으면서도 같이 있을 때 재밌고 듬직해 자꾸 술친구 하게 됐던 의종오빠, 짓궂은 장난도 받아주고 뒷자리에서 친절하게

저의 구글까지 되어줬던 지훈오빠, 겉으로 툴툴거려도 맘은 따뜻해 대화하는 게 즐거웠던 종민이, 옆자리에서 제 기분 컨트롤 해주느라 고생한 눈치 빠른 호연이, 후배를 대하는 법을 고민할 수 있게 해준 센스 있고 예쁜 혜란이, 조용하지만 똑 부러지는 모습에 의지하게 됐었던 지민이, 굳은 실험실 일도 마다하지 않고 웃으면서 해준 사랑꾼 성호까지 그동안 저 때문에 고생 많았다고, 고마웠다고 전하고 싶습니다.

대학원 12학번 동기들이라는 단어만으로 힘이 되어주었던, 시크한 듯 따뜻하게 챙겨줬던 형안이, 만날 때마다 에너지를 선물해줬던 대학원 유일한 여자 동기 채윤이, 멍했다가 다정했다가 재밌었다가 하던 다솔이, 내 생각을 끝까지 들어줬던 오빠 같은 동생 승언이, 웨딩 컨설턴트가 되어줬던 자주 보지 못해 아쉬운 현호오빠에게도 고맙습니다. 학부 때부터 옆에서 조언을 많이 해줬던 훈영이, 특이하지만 따뜻한 마음씨로 챙겨줬던 승규오빠, 대학생활 시작의 은인이 되어준 왕호, 직설적이지만 마음은 따뜻한 태환오빠, 대학교 시절부터 계속 옆에서 재밌게 똑똑하게 사는 법을 말해주는 용철오빠 모두 감사합니다. 그리고 만날 때마다 너무 잘 통해서 깜짝깜짝 놀라는 우리 이쁜이들, 대학생 때부터 친구처럼 뭐든 응원해준 민희, 항상 예쁜 마음씨가 보이는 서연이, 씩씩한 여장부 스타일 지현이, 기 센 언니들 사이에서도 잘 버텨주는 착한 동생 지영이 모두 고맙습니다. 제 행동과 말투 하나하나 먼저 읽고 감싸줬던 똑똑한 망고오빠, 학부 때부터 친한 선배이자 재밌는 친구 같았던 우석오빠, 미국으로 연구하러 가서도 한국에서 그랬던 것처럼 함께

대화하는 게 재밌고 고마운 두성오빠, 왜 대기과 마당발인지 알 수 있을 정도로 소소한 일까지 알아줬던 창현이, 시크하지만 따뜻한 언니 같은 진주, 같이 대학원 생활의 고충을 공유하던 동생 강현이, 조용히 먼저 다가와서 챙겨줬던 유나, 느낌과 실력을 겸비한 작가님 재연이, 두 오빠가 떠난 빈자리를 따뜻하게 채워 주신 준우오빠, 저의 수다 방출구가 되어 주신 지영언니, 옆 실험실 선배여도 조언을 아끼지 않으셨던 민중오빠, 먼저 다가와 챙겨 주신 정화언니 모두 감사합니다. 그리고 영국 기상청 연수 생활 때 안팎으로 도움을 주신 은정언니와 기상청에서 모델을 수행할 수 있게 많은 도움을 주신 김미자 연구사님께도 감사의 인사를 드립니다.

어쩌면 연구적으로 만난 인연보다 더 소중하고 고마운 친구들에게도 지금까지 잘 해낼 수 있게 정신적, 정서적으로 도와줘서 고맙다는 말을 전하고 싶습니다. 지금까지 서로에게 큰 힘이 되어주었고, 앞으로도 평생 서로 힘이 되어줄 것이라 믿어 의심치 않는 사랑하는 07 깔녀들, 소심녀 방실이, 개그녀 상희, 비염녀 민정이, 집착녀 명이, 18년 동안 기센 친구 옆에서 착하게 함께 해주는 단짝친구 은지, 한국 · 일본 · 영국 · 미국 어디에서도 안정감을 주었던, 그럼에도 저는 그만큼 하지 못한 것 같아 미안한 효정이 정말 고맙습니다. 그리고 대학원 기숙사 룸메이트로 만나 이제는 둘도 없는 친구이자 서포터즈가 된 흔정이와 그 옆에서 함께 해주는 앤디, 영국에서 서로 의지하면서 친해져 한국에서도 그 소중한 인연을 이어주는 수진이, 항상 이박사님이라고 불러주면서 제 자존감을 올려주는

상아 선생님, 인생에서 최고 예민했던 고등학교 시절부터 나와 친구 해준 착한 친구들 주경이, 세정이, 묘선이, 새롬이, 언제 만나도 편하고 재밌는 친구 하영이, 요즘 고등학교 때처럼 다시 친해지면서 즐거움을 주는 민희, 윤정이, 우리, 남자친구의 친구로 시작된 인연이지만 지금은 진짜 내 친구 같은 태수 선생님, 대학원 시절 내내 함께 생각을 공유해준 멋진 교수님 세은언니, 주기적으로 고민 들어주고 힐링을 선물해준 네이처네일 한다연 쌤에게도 감사합니다.

지금도 하늘나라에서 저를 가장 예쁘다고 해주실 우리 할아버지, 항상 주시는 사랑에 비해 자주 찾아뵙지 못해 죄송한 우리 할머니, 편찮으시지만 오래오래 저희와 함께해주셨으면 하는 외할아버지, 큰소리치시지만 가장 마음 여리신 외할머니, 그리고 첫 번째 조카라고 뭐든 항상 조건 없이 주시는 승하 삼촌, 승한이 삼촌, 영택이 삼촌, 정택이 삼촌 고맙습니다. 어렸을 적부터 친구처럼 지내며 지금까지 항상 응원해준 사촌 동생 연주, 근영이, 그리고 채은이를 포함한, 제가 이 세상에 존재할 수 있게, 그리고 제가 존재함에 기뻐해주신 친가, 외가 친척분들께도 감사드립니다.

며느리 연구에 대해 먼저 궁금해 해주시는 아버님, 연구하느라 고생한다며 매번 맛있는 음식 챙겨 주시는 어머님, 이벤트가 있을 때마다 따뜻하게 먼저 챙겨 주시는 큰언니, 큰매형과 힘들 때마다 날아온 영상 속 힐링제가 되어준 귀여운 솔이, 항상 물질적으로나 정신적으로 적극 서포트해주시는 작은언니, 작은매형을 포함한 이제는 새로운 가족이 된

분들께도 감사하단 말씀드리고 싶습니다.

마지막으로 제가 세상에서 가장 사랑하는 우리 가족에게 정말 많이 고맙다는 말을 전하고 싶습니다. 제가 박사가 된 것을 누구보다 기뻐해주시고 자랑스럽게 생각해주시는 사랑하는 우리 아빠, 어렸을 때부터 욕심 많은 딸이 원하는 것은 다 해주려고 노력해주신 사랑하는 우리 엄마, 33년간 조건 없이 주신 사랑에 보답할 수 있게 앞으로 더 멋진 큰딸이 되겠습니다. 친구처럼 때로는 언니처럼 조언해주고, 지금은 자랑스러운 한 회사의 사장님으로서 물질적으로도 서포트 해주는 하나뿐인 동생 다미에게 언니가 공부할 동안 정신적으로 힘이 되어줘서 정말 고마웠다고, 앞으로도 지금처럼만 세상에 둘도 없는 자매로 잘 지내보자고 말하고 싶습니다.

그리고 제가 가장 힘들었던 시기에 작은 일에도 예민하게 반응할 때마다 큰 나무처럼 넓은 그늘을 만들어주고 그 안에서 힐링할 수 있게 해준, 5년간의 연애 끝에 2020년 5월부로 제 남편이 되어준, 충회에게도 그동안 저 때문에 고생 많았다고, 제일 많이 고마웠다고 말해주고 싶습니다. 앞으로 더 사랑스러운 와이프가 되겠습니다, 사랑합니다.